

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 09:48:15 ; Search time 20 seconds
(without alignments)
2193.167 Million cell updates/sec

Title: US-09-905-589A-2
Perfect score: 2364
Sequence: 1 MKTISNGSLRVAKVAYPLG.....ALGAIFYIDSLNRQKSPAS 456

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	590.5	25.0	479	T23508	hypothetical prote
2	508	21.5	568	T33109	probable guanine
3	498	21.1	518	A40732	guanosine-diphosph
4	485	20.5	454	JC4616	aprase (EC 3.6.1.
5	442.5	18.7	455	S48859	nucleoside triphos
6	372	15.7	516	G84442	probable nucleosid
7	343	14.5	485	T34147	hypothetical prote
8	339.5	14.4	483	D86276	hypothetical prote
9	329.5	13.9	630	S50463	hypothetical prote
10	307	13.0	557	T16696	hypothetical prote
11	301	12.7	405	E86276	hypothetical prote
12	285.5	12.1	510	I56242	lymphoid cell acti
13	281.5	11.9	572	T40856	probable nucleotid
14	245.5	10.4	1052	T04439	hypothetical prote
15	219	9.3	508	C86276	7A.9.33 protein -
16	140.5	5.9	628	A55421	nucleoside-triphos
17	111	4.7	679	C71007	probable formate d
18	108.5	4.6	1186	T12737	tail protein - Met
19	106.5	4.5	665	E75124	formate dehydrogen
20	104.5	4.4	322	T24948	hypothetical prote
21	103.5	4.4	322	T27308	hypothetical prote
22	102.5	4.3	440	I39847	cell-division prot
23	102	4.3	475	G72274	glucamyl tRNA-Gln
24	102	4.3	623	G81346	heat shock protein
25	102	4.3	4162	T42633	connectin/titin -
26	100	4.2	1233	T37045	nitrate reductase
27	100	4.2	26926	I38344	titin, cardiac mus
28	99.5	4.2	470	I31339	dihydrolipoamide d
29	99.5	4.2	504	S17724	H+-transporting tw

30 99.5 4.2 633 2 I39585
31 99.5 4.2 2411 2 A46299
32 98.5 4.2 633 2 AE2591
33 98.5 4.2 633 2 E97373
34 97.5 4.1 415 1 C64473
35 97.5 4.1 548 2 S32615
36 97 4.1 2511 2 T14591
37 96.5 4.1 515 2 T35970
38 96 4.1 711 2 S73898
39 96 4.1 1217 2 T25894
40 95.5 4.0 325 2 T32940
41 95.5 4.0 325 2 D87733
42 95 4.0 527 1 S25478
43 95 4.0 549 2 S23448
44 95 4.0 15281 2 S41309
45 94.5 4.0 489 2 F97119

ALIGNMENTS

RESULT 1
T23508
hypothetical protein K08H10.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T23508
R:Gardner, A.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19750
A:Accession: T23508
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-479 <WIL>
A:Cross-references: EMBL:Z83113; PIDN:CAB05544.1; GSPDB:GN00023; CESP:K08H10.4
C:Genetics:
A:Gene: CESP:K08H10.4
A:Map position: 5
A:Introns: 36/3; 83/3; 189/1; 300/2; 412/3
C:Superfamily: nucleoside triphosphatase chromatin-associated

Query Match 25.0%; Score 590.5; DB 2; Length 479;
Best Local Similarity 34.2%; Pred. No. 2.1e-40;
Matches 134; Conservative 78; Mismatches 155; Indels 25; Gaps 9;
Qy 73 FYGIMEDAGSTGRVHVFOFTPRPRT-----TLTHETFKAVKPGISAYADDDVKSQAQ 127
Db 42 FFIIVIDAGSTGRVHVFOFTPRPRT-----TLTHETFKAVKPGISAYADDDVKSQAQ 101
Qy 128 IRELLDVAQDIPDFWKATPLVLKATAGLRLPGEKAQKLQKVE-VFKASPFVLGDD 186
Db 102 LEPLQARKEVPHFMWEKTPITLKATAGLRLPGEKAQKLQKVE-VFKASPFVLGDD 161
Qy 187 CVSIMNGTGVSAWITINFLTGLS-----KTPGGSSVGMIDLGCGSTQIAFLPRV 237
Db 162 AVNMPGSGGVSWFTLNILLETIFDETVGKPAHRSVAADFLLGGSTQLTYPNN 221
Qy 238 EGTLLQASPPGYLTALRMFNRTKLYSYSLGLGLMSARLAILGVEGQPAKDGKELVSPC 297
Db 222 EAVP-SEHVGVERDIDFFGHIRLFTSHFLNGLIAARLNLQLETDNE-ESTHQLTSC 280
Qy 298 LSPSEK-GEWEHARTVRSQKAAASLHELCAARVSEVLQ-NRVHRTVEEVKGVDFVAFS 355
Db 281 MPEGYQLTEWEYALKFWNING--SSSHSFESCVYGTNNFVSESEIMELRELKSGPVLFPS 338
Qy 356 YYDLAAGVGLIDAEKGGSLVVGDFEFAAKYVCVLTETQFQSS----PFSCMDLTYY-SL 410
Db 339 YFFPRALNSGLVKNEGKIELRQFKEAETACRREXTEIDDSGSHWMPWQCLDLYIVSL 398
Qy 411 LLOEFGPPRSKVLKLRKIDNVETSWALGALF 442
Db 399 LRDGVOQFEDNQLVLAKKIKGMEVSWGQGLAF 430

C:Accession: JC4616; PC4147
R:Handa, M.; Guidotti, G.
Biochem. Biophys. Res. Commun. 218, 916-923, 1996
A:Title: Purification and cloning of a soluble ATP-diphosphohydrolase (Apyrase) from pea
A:Reference number: JC4616; MUID:96158985; PMID:8579614
A:Accession: JC4616
A:Molecule type: mRNA
A:Residues: 1-454 <HAN>
A:Cross-references: GB:U58597; NID:g1381632; PIDN:AAB02720.1; PID:g1381633
A:Accession: PC4147
A:Molecule type: protein
A:Residues: 59-95;96-131;132-160;236-253;332-345 <HA2>
A:Experimental source: tubers
A:Note: The authors translated the codon GCA for residue 215 as Gly
C:Comment: This enzyme belongs to a family of E-type Atpases, and it catalyzes the hydrolysis of ATP to ADP and inorganic phosphate. It has a high affinity for ATP and is involved in starch synthesis.
C:Genetics:
A:Gene: trop1
C:Superfamily: nucleoside triphosphatase chromatin-associated
C:Keywords: glycoprotein; hydrolase; transmembrane protein
F:1-30/Domain: signal sequence #status predicted <SIG>
F:8-25/Domain: transmembrane #status predicted <TM>
F:31-454/Product: ATP-diphosphohydrolase #status predicted <MAT>
F:44-65/Region: actin-heat shock protein 70-hexokinase beta-phosphate binding
F:44-65/Region: nucleotide binding #status predicted
F:192-212/Region: actin-heat shock protein 70-hexokinase gamma-phosphate binding
F:192-212/Region: nucleotide binding #status predicted
F:390-410,427-446/Region: hydrophobic carboxyl end
F:151,262/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 20.5%; Score 485; DB 2; Length 454;
Best Local Similarity 31.9%; Pred. No. 8.8e-32;
Matches 138; Conservative 68; Mismatches 183; Indels 44; Gaps 12;

QY 59 AHSPLGTAADGHE-VFYGMFMAGSTGTRVHVQFQTRPPREPTLTH-ETFKAVKPGLSA 116
DB 30 AQPLRRLLSHSEHYAVIFDAGTSGSRVHVFRDEKLGILPIGNNIEYFATEPGLSS 89
QY 117 YADDVEKSAQGTRELLDVAQDIPDFWKATPLVLKATAGLLPGEAKQLQKVEVF 176
DB 90 YAEADPKAAANSLFPLDGAEGVVPQELQSETPLELGATAGLRLKGDAAEKLQAVNLV 149
QY 177 K-ASPLVGDCCYSIMNGTDEGVSATITNFLTSGKTPGSSVGMGLDGGSGSTQIAFLP 235
DB 150 KNOSTFHSKDQWVTILDQEGSYMAALNLLGNLKGDKYSTATTIDLGGSVQVAYAI 209
QY 236 RVEGTLQA-----SPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGK 291
DB 210 SNEQFAKAPQNEDEPVPQKHLMSKDXNLYVHSYLYNGQLAGRAEIF-----KASR 261
QY 292 ELVSPCLSPSPFKGEWEHAETVYRVSGQKAAASHLHCAARVSEVLQ-NRVHRTSEVK--- 347
DB 262 NESNPCALEGCDGYGYGGVDYKVPKPKGSS-WKRCRLTRHALKINAKCNIECTFNG 320
QY 348 -----HVDFYAFSYVYDLAAGVGLIDAEGKSLVWG-DFEIAAKVYVCR----- 390
DB 321 VNWGGGGDGGKXNTHASFFDIDGAGVIVDTKFPALAKPIQYLNAAKACQINVADIKS 380
QY 391 --LETQPSQSPFCMDLTY-VSLLQEFPGFPRSKVLKTRKIDN-----VETSWALGA 443
DB 381 IFPKTDNRNIPYLCMDLIYEYTLVDGFLGNPKHEITVHDVQYKNLYVGAANPLGCAID 440
QY 444 YIDSLNRQKSPAS 456
DB 441 LVSSSTTKIRVAS 453

RESULT 5
S48859
nucleoside triphosphatase precursor, chromatin-associated - garden pea
C:Species: Pisum sativum (garden pea)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 29-Sep-1999

C:Accession: S65147; S48859
R:Heieh, H.L.; Tong, C.G.; Thomas, C.; Roux, S.J.
Plant Mol. Biol. 30, 135-147, 1996
A:Title: Light-modulated abundance of an mRNA encoding a calmodulin-regulated, chromatin-associated triphosphatase
A:Reference number: S65141; MUID:96197404; PMID:8616230
A:Accession: S65147
A:Molecule type: mRNA
A:Residues: 1-455 <HS2>
A:Cross-references: EMBL:Z32743; NID:g563611; PIDN:CAA83655.1; PID:g563612
A:Superfamily: nucleoside triphosphatase chromatin-associated
C:Keywords: nucleus

Query Match 18.7%; Score 442.5; DB 2; Length 455;
Best Local Similarity 29.2%; Pred. No. 2.7e-28;
Matches 125; Conservative 79; Mismatches 145; Indels 79; Gaps 15;

QY 74 YGIMFDAGSTGTRVHVQFQTRPPREPTLTH-----ETFKAVKPGLSAYADDVEKSAQGI 128
DB 44 YAVFDAGSTGSRHVHF-----NQNLDDLHLHGKGVYNNKITPGLSSVYANNPEQAASL 99
QY 129 REILLVAKQDIPDFWKATPLVLKATAGLLPGEAKQLQKVEVF-KASPLVGDCC 187
DB 100 IPLLQAEADVVPDLOPKTEPRLGATAGLLNGDASEKILQSVDRMLSNRSTFNVPQDA 159
QY 188 VSIMNGTDEGVSATITNFLTSGKTPGSSVGMGLDGGSGSTQIAF-----LPRVE 238
DB 160 VSIIDGQEGSYLWTVVYALGNLKKYTVGVVILGGSSVQMAVAVSKTAKNAPKV- 218
QY 239 GTLQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCL 298
DB 219 --ADGDDP-VIKKVLKGIPIYDLYVHSYLFHFGREASRAEILKLTFRSP-----NPCL 267
QY 299 SPSPKGEWEHAETVYRVSGQKAAASHLHCAARVSEVLQNRVHRTSEVKH----- 348
DB 268 LAGFNGIYTSGBEFKATATVSGANFNK-----CKTIRKALKLNYPQYQNTFG 318
QY 349 -----VDFYAFSYVYDLAAGVGLIDAEGKSLV-VGDFFEIAAKVYVCR-TLETQP 395
DB 319 GIWNGGGGQKQLFASSTFPYLPEDTGMVDASTPNFILRPVDITETKAEACALNPFDAK 378
QY 396 QSSPF-----SCMDLTVSLLQEFPGFPRSKVLKLT--RKIDN-----VETSWALGA 440
DB 379 STYFPLDKKNVASYVCMDLIYQVYLLVD-GFGLDPLQKITSGKEIYIQDAIVEAANPLGN 437
QY 441 IFHYIDSL 448
DB 438 AVEAISAL 445

RESULT 6
G84442
probable nucleoside triphosphatase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 18-Nov-2002
C:Accession: G84442
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mili, X.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.; Koo, H.; Moffat, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: G84442
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-516 <STO>
A:Cross-references: GB:AS002093; NID:g3461821; PIDN:AAC32915.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g02970
A:Map position: 2
C:Superfamily: nucleoside triphosphatase chromatin-associated

Query Match 15.7%; Score 372; DB 2; Length 516;
Best Local Similarity 29.0%; Pred. No. 1.9e-22;

Matches 113; Conservative 67; Mismatches 173; Indels 36; Gaps 13;

QY 74 YGIMPDAGSTGTRVAVFOFT---RPRETPTLTHETFKAVKPGLSAVADVDEKSAQGR 129
 Db 68 YSVVIDGSGTGRIRHVGRISGKPVFEFRGANVYASLK-LHPGLSAFADDPDGAHSVSLT 126

QY 130 EILDVAKODIPDFPKATPLVLKATAGRLLPGEKQKLLQKVEFKASPFVGDGCS 189
 Db 127 ELVEFAKGRVPGKGMITEVRLMATAQMLLEIPVQEKILGVARRVLYKSGFLFDEMAS 186

QY 190 INMGDEGVASAMITINFLTSGLSKTPGSSVGMULDGGSGTQIAFLPRVGTQLQASDPGYL 249
 Db 187 VTSQSDGVAVMYAVANPALGSLGDPFKTTGIVELGASAQVTFV-----SSEPMPEPS 241

QY 250 TALRMFRTYKLYSYSLGSLMSARLAIIGVEG-----QPADGKELVSPCLSPSF 302
 Db 242 RTISFGVNTYLYSHSFHFGQNAADKMLGSLSDHNSAVPEPTR-KIFTPCAPKGY 300

QY 303 K--GEWEH-----AEVTVRSGQKAASLHELCAARVSEVLQNRVHRETEVGVDFYAF 354
 Db 301 NLDAATQKHLGSLAELS-RLSDSFGKGNYSQCRSAALTLIDGNGRI---LIIAGP 355

QY 355 SY-YIDLAAGVGLDAEKGSLLVGDPELIAKYVCTLETPQSPSCMDLTY-VSLLL 412
 Db 356 SLEFFGLGKRWALSNMISAGERFCG--EDWSKLRVDPSLHEBDLTRYCFSSAYIVSLH 413

QY 413 QEGFEP-RSKVLLKTRKIDNVETSMALGA 440
 Db 414 DTGLPDDDERIGYANAGDIPDMALGA 442

RESULT 7
 T34147
 hypothetical protein C33H5.14 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000

C:Accession: T34147
 R:Bradshaw, H.; Stellyes, L.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid C33H5.
 A:Reference number: Z21482
 A:Accession: T34147
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-485

 A:Cross-references: EMBL:U41007; PIDN:AAA82272.1; CESP:C33H5.14
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:C33H5.14
 A:Introns: 22/1; 83/1; 120/2; 167/1; 269/3; 399/3
 C:Superfamily: nucleoside triphosphatase chromatin-associated

Query Match 14.5%; Score 343; DB 2; Length 485;
 Best Local Similarity 25.8%; Pred. No. 4.2e-20;
 Matches 113; Conservative 74; Mismatches 157; Indels 94; Gaps 16;

QY 70 HEVFGIMPDAGSTGTRVAVFOFT---RPRETPTLTHET---FKAVKPGLSAVADV 121
 Db 21 NNKYGIVICDAGSSGTR--LFVYTLKPLSGGLTIDTLIHSESPVAVKVPGLSSPDKP 78

QY 122 EKSAQIGRELIDVAKODIPDFPKATPLVLKATAGRLLPGEKQKLLQKVEFKASPF 181
 Db 79 EGQVAVYLPLRLFAEHIPEGLGETDILLPATGMRLLPEAQDAIILKLNQNLKSVTA 138

QY 182 L-VGDDCVSLMGDTGVSAMITINFLTSGLSKTPGSSVGMULDGGSGTQIAFLPRVGT 240
 Db 139 LEVSPSNRIIDGAEIGYSWIAVNYIIGRPDKENDSTYGMIDMGGSVQJAF--ETANE 196

QY 241 LQASDPGYLTALRM-----FNRKYKLYSYSLGSLMSARLAIIGVEGQPADGKELY 294
 Db 197 KESYNGAVYELNLSIENEDYKTKIYITTLGSGAEG---LKKYENLVVSGNS-- 250

QY 295 SFKGEWHAELVTVRSGQKAASLHELCAARVSEV----- 335

Db 251 NDSCSPGRLRLIEFTVNGGEW-----DVCLAQVSSLIDGKAQ 291

QY 336 -----LONRVHTEBVKHVDFAFSYYDLAAGVGLDAEKGSLLVGDPELIAKY 387
 Db 292 SCNPPTCELRNVIAPSVSLSTVQLYGFSEYWTYSNG-----SGGEYHQKFTDEVKRY 346

QY 388 CR-----TLETPQSPF-----SCMDLTVSLLQFPPRSGVLTTR--KID 430
 Db 347 CQKMDIQQGFKQNEFPNADIELGNCFRGAANVTSLVD-GRVNDKTHLPQSVLKIA 405

QY 431 NVETSMALGALFHYIDSL 448
 Db 406 GEEMQWALGAMLYHSKDL 423

RESULT 8
 D86276
 hypothetical protein F7A19.34 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 18-Nov-2002

C:Accession: D86276
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federapfel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chinn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Hulzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzilli, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Xu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A6141; MUID:21016719; PMID:11130712
 A:Accession: D86276
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-483 <STO>
 A:Cross-references: GB:AE005172; NID:95080801; PIDN:AAA39311.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 A:Superfamily: nucleoside triphosphatase chromatin-associated

Query Match 14.4%; Score 339.5; DB 2; Length 483;
 Best Local Similarity 26.5%; Pred. No. 8.1e-20;
 Matches 108; Conservative 78; Mismatches 166; Indels 55; Gaps 13;

QY 74 YGIMPDAGSTGTRVAVFOFT---RPRETPTLTHETFKAVKPGLSAVADVDEKSAQGR 129
 Db 70 YSVLIDAGSGSTGTRVAVFGWFSGRKVPDFGKHANLK-LTPGLSSYADNDEGASVSVT 128

QY 130 EILDVAKODIPDFPKATPLVLKATAGRLLPGEKQKLLQKVEFKASPFVGDGCS 189
 Db 129 KLYEFAKORIPKMRPRSDIRLMATAGRLLEVPQQLVETRVVLSGGMFDEMAN 188

QY 190 INMGDEGVASAMITINFLTSGLSKTPGSSVGMULDGGSGTQIAFLPRVGTQLQASDPGYL 249
 Db 189 VTSQSDGVAVMYAVANPALGSLGDPFKTTGIVELGASAQVTFV-----SSEHPPEPS 243

QY 250 TALRMFRTYKLYSYSLGSLMSARLAIIGVEGQPADGKELY 294
 Db 244 RTIAYGNISYITYSHSFIDYGDAAALKLELQNSANSTVDGVVEDPCTPKGYIYDTNS 303

QY 291 KELVSPCL--DSFKGEWEH-----EVTYRSGQKAASLHELCA--AEVSEVLQNR 339
 Db 304 KYSSGFTLADBSKAGSLQAAGNFKCSATFALKKEGENCELYHSGSGSFTFDLQGS 363

QY 340 VERTEEVXVDVFAFSYYDLAAGVGLDAEKGSLLVGDPELIAKYVCTLETPQSPSF 399
 Db 364 FLAT-----ASFYAKAPFELEKGMVLSLIPAGRRYCG--EWSKLLIEYFTDEEYLR 416

QY 400 FSCMDLTY-VSLLLQFPPR-----SKVLLKTRKIDNVETSMALGA 440
 Db 417 GYFSAAYTISVHLHSLGIALDDESTITYASKGK--HIPDMALGA 461

RESULT 9

S50463
hypothetical protein YER005w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 19-Apr-2002
C:Accession: S50463
R:Dieterich, F.S.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmid 9537, 9581, 9495, 9867, and lambda
A:Reference number: S50433
A:Accession: S50463
A:Molecule type: DNA
A:Residues: 1-630 <DIE>
A:Cross-references: EMBL:U18778; NID:g603592; PID:g603597; GSPDB:GN00005; MIPS:YER005w
C:Genetics:
A:Gene: SGD:YND1; MIPS:YER005w
A:Cross-references: SGD:S0000807
A:Map position: 5R

Query Match 13.9%; Score 329.5; DB 2; Length 630;
Best Local Similarity 26.9%; Pred. No. 7.8e-19;
Matches 122; Conservative 72; Mismatches 163; Indels 97; Gaps 22;

Qy 74 YGIMFDAGSTGRVHVQFTRPREPTLTHE-----TPKAVKP 112
Db 10 FGIVIDAGSGSRIRHVFKV---QDTESLLHATNQDSQSILOVSPHIHQEKDWTFK-LNP 64
Qy 113 GLSAYADDDVEKSAQG---IRELLDVAKODIPDFWKATPLVKATAGLRLPGEKAQK 167
Db 65 GLSSP---EKPDQAYKSHIKPLDFAKNIIPESHNSCFVIOATAGMRLLPQDIQSS 120
Qy 168 LIQKQKVEFK-ASPFIVGDDC---VSMNGTDEGVSAMTINFLGSLK--TPGGS---S 218
Db 121 ILDGLCQGLKHPAEFLV-EDCSAQIQVIDCETEGLYGMLNLYGHFNVDYNPEVSDHFT 179
Qy 219 VGMLDLGGSTQIAFLPRVEGLQASPPGYLTA-LEMEN---RTYKLYSYSVLGLGMSA 274
Db 180 FGMMDGGASTQIAFAHDSGEIARHDDIATIFRSVNGDIQKWDVFSVTLGFGANQA 239
Qy 275 R---LAILGGVEGQAKD-----GKELVSPCLSPSPKGEWEHAEVTVRVSG-----OKA 320
Db 240 RRYLAQLNLTLPENTNDYENDDFSTRNLNDFCMRPGSSTDPEFKDTIFHIAAGSGNVEQC 299
Qy 321 AASLHELCAARV-----SEVLQRVHTE-EVQHVDFAPSYIYDLAAGVLIDAEKGSGL 375
Db 300 TRSIYPLLLKNWPCDDPECLFNGVHAPRIDFANDKFIGTSEWYTANDV---FKLGGEY 355
Qy 376 VVGDFEIAKYVCTRLETO-----POS-SPFSCMDLTVYSLLLQOE-FGFFRS 420
Db 356 NFDKFSKSLREFNSNWTQILANSKGVYNSIPENFLKDACFKGNVNLILHEGDFMPEI 415
Qy 421 KV-----LKLPRKIDNVETSWALGAIFY 444
Db 416 DVDAENVNDRPLFQSVKEVEERELSWTLGRILLY 449

RESULT 10

T16696
hypothetical protein R07E4.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16696
R:Miller, N.

submitted to the EMBL Data Library, October 1995
A:Description: The sequence of C. elegans cosmid R07E4.
A:Reference number: Z18561
A:Accession: T16696
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-557 <ML>
A:Cross-references: EMBL:U39652; NID:g1049390; PID:g1049394; PIDN:AA80403.1; CESP:R07E4

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:R07E4.4

A:Introns: 39/1; 67/1; 145/2; 192/1; 244/3; 317/3; 451/3; 488/2

Query Match

13.0%; Score 307; DB 2; Length 557;

Best Local Similarity 23.5%; Pred. No. 4.6e-17;

Matches 119; Conservative 88; Mismatches 178; Indels 122; Gaps 23;

Qy 13 AKVAYPLGLCVGVFIYVAVIKMHRATATCAFFSITRAAFGARWGQQAHPGLGTAADGHEV 72
Db 14 AMIFPP-----VIVFIYV-----EAHTSPKIADQER 42
Qy 73 FYGIMFDAGSTGRVHVQFTRPP-----RETPTL--THETFKAVPGLSAYADDDVEKSA 125
Db 43 SYGVICDAGSTGRVFIYVAVIKMHRATATCAFFSITRAAFGARWGQQAHPGLGTAADGHEV 102
Qy 126 OGIRELLDVAKODIPDFWKATPLVKATAGLRLP-----CEKAQKLLQKVEKFKAS 179
Db 103 EYLRPLMELAEHPIPEEKRYTFVFFATAGMRLIDPEYVLQKAEVLEKLEKLPKIT 162
Qy 180 PFLVGGDDCVSIMGTDGVSAMTINFL-----TGLSKTPGGS-----SVGMLDLGG 226
Db 163 SMQVLKHEHRIIEGKWEIYSIAVAVNVALGKFNKTATLOPPGTSPAHARQKTVGMIDMG 222
Qy 227 GSTQIAP-LPRVEGTLCASPPGYLTAIR---MENRTYKLYSYSVLGL-----GLMSARLA 277
Db 223 ASAQIAPELPTDPSFNSINVENINLGRDDSLFK--YKLFVTTFILOYGVNEGIRKYEHM 280
Qy 278 ILGVEGQAPKDGKELVSPCLSPSPKGEWEHAEVTVRVSGQ---KAAASLHELCAARVSE 334
Db 281 LLSKLKQD---NGTVIQDDCMPLNL-----HKTVTLE-NGENFVRRTGNTCSNEVKK 331
Qy 335 VL-----QNRVHRTVEVK-----HYDFFAFS-YYDLAGVGLIDAEKGSGL 375
Db 332 LLNPESSEYCKAEAAKCYFAGVAPAPSPILPSNIEMFYGFSEYWTHTDVLGL-----GGQY 386
Qy 376 VVGDFEIAKYVCTRLETO-----RTELETOFQSSPF-----SCMDLTVYSLLLQBFQPRSK 421
Db 387 DAENIAKKTQYCSKRWSTQASKKLYPRADDERLRTQCFKSAMITSVLHD-GFSVDK 445
Qy 422 VLKTRKIDNV---EFTSWALGAIFYI 445
Db 446 THNKFQSVSTIAGQEVQWALGAMIYHM 472

RESULT 11

E86276
hypothetical protein Fl1417.1 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001

C:Accession: E86276

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

A:Authors: Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: E86276

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-405 <STO>

A:Cross-references: GB:AE005172; NID:g7626666; PIDN:AAF43924.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 12.7%; Score 301; DB 2; Length 405;

Best Local Similarity 25.4%; Pred. No. 9.1e-17;

C/Accession: T04439
 R/By: M.; Bene, V.; Rechmann, S.; Borikova, D.; Ansoorge, W.; Bancroft, I.; Mewes, H.W.
 Submitted to the Protein Sequence Database, April 1998
 A/Reference number: 215359
 A/Accession: T04439
 A/Molecule type: DNA
 A/Residues: 1-1052 <SEV>
 A/Cross-references: EMBL:AL021687
 A/Experimental source: cultivar Columbia; BAC clone T18B16
 C/Genetics:
 A:Map position: 4
 A:Insertions: 79/3; 118/3; 217/1; 295/3; 396/3; 531/2; 815/3
 A:Note: T18B16.150

Query Match 10.4%; Score 245.5; DB 2; Length 1052;
 Best Local Similarity 25.3%; Pred. No. 1.2e-11;
 Matches 116; Conservative 60; Mismatches 185; Indels 97; Gaps 20;

QY 55 MGQQAHS-----PCTAADGHE-----VYGMFDAGSTGTVHPQFT 93
 Db 483 WQNGSSFSKESKGVPMGTPTSTRKLIKAVMTWCLFLPAFLVYISITGRAYVYOAS 542
 QY 94 ---RPREPTLTETETKAV-----KPLSAVADDEKSAQIGRELIDVAK 136
 Db 543 INYKKSLSPLVMSLTREGISRSRGRAYDMETEPFDLVNRRGLKTAIKELIOWAE 602
 QY 137 QDLPFDWKATPLVLAATAGLRLLPGEKQKLOKVEKFAKPFVLDVDCVSIINGTDE 196
 Db 603 KQIPKNAHRTSLFVYATAGVRLRPADESSWILGNVSIKASPFTRRRRWVXIISTEE 662
 QY 197 GVSAMITINFLTGLSL--TPGGSSVGMDLGGSGSTQIAFLPRVEGTLOASPGVLTLMRF 255
 Db 663 AYPFWTALNYOTSMGLPKPKATFGALDLGGSSIQVTF---EKERTHETNIN--IRIG 717
 QY 256 NRTYKLYSYSLGLMSA-----RLAILGVYEGGPANDGK--ELVSPCLSPSEKEME 307
 Db 718 SVNHLTAVSYLAGLMDAFDRSVVHLKLLPNVNSKDLIEGLKEMHGPCLNGYNGQYI 777
 QY 308 HAEVTYVSQOKAASLHEL-----CAARVSE--VLONRVTRTEVGHVDFYAF 354
 Db 778 CSQCASVQGGKKKSGVSIKLVGAPNMGESALAKKAPCALPGVPR---DHGGFYAV 833
 QY 355 S-----YYVDIAGVGLIDA--EKGSILVGDFFELIAKYVCTLETPQSSPF---SCMD 404
 Db 834 SGFVYVRRPFVLSABASLDVLEKG-----REFCDKAMQVART--SVSPQ--PFIEGYCFR 885
 QY 405 LTVYSLLIOGFPRSKVLTUKTK---LDNVTSMALG 439
 Db 886 APYTVSLIR-----GLYITDKQIITIGSGSITWTLG 916

RESULT 15

C86276
 7A19.33 protein - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 18-Nov-2002
 C/Accession: C86276
 R/Theologian, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 anson, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000
 A/Authors: Hunter, U.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Ventier, J.C.; Davis, R.W.
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: A86141; MUID:21016719; PMID:11130712
 A/Accession: C86276
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-508 <STO>
 A/Cross-references: GB:AE005172; NID:G5080800; PIDN:AAD39310.1; GSPDB:GN00141

C/Genetics:
 A:Map position: 1
 C/Superfamily: nucleoside triphosphatase chromatin-associated

Query Match 9.3%; Score 219; DB 2; Length 508;
 Best Local Similarity 22.3%; Pred. No. 6.6e-10;
 Matches 109; Conservative 83; Mismatches 196; Indels 100; Gaps 21;

QY 9 SLRAKVAVPLGLCVGFVYVAVIKWHATATQAFSTIRAPAGARWGQAHSPLGTAD 68
 Db 26 SKRTKSIIFVIVACTTALGLFLG-----YSLRGKRR----- 61
 QY 69 GHEVFYGMFDAGSTGTVHPQFTPTLT--HETKAVK--PGLSAVADDEKRS 124
 Db 62 -VSLHYSVLIIDGSSGSTRVHVGY--RIESGKVPFDGGEENYASLKLSPGLSAVADNPEGV 119
 QY 125 AGGIRELDVAKODIPFPWKATPLVLAATAGLRLLPGEKQKLOKVEKFAKPFVLDVDCVSIINGTDE 184
 Db 120 SESVTELVFPAKGRVHKGLKKSQDRLMAYTGMRLBELVPOQLDVTTRVLRSSGFDPR 179
 QY 185 DDCVSIINGTDEGVAMITINFLTGLSLTTPGGSSVGMDLGGSGSTQIAFLPRVEGTLOAS 244
 Db 180 DEMASVISIILENPDQIMKV--YMLGLL-----LIMSVRLEVFV-----STELV 223
 QY 245 PRGYLTALMFRRTKLYSYSLGLMSARLAILGVYEGGPANDGK--ELVSPCLSPSEKEME 299
 Db 224 PSFPRTLAYGNVSYNLVSHSFLDFGDDAAQEKLSLYNSAANSTGIVPDDCIPKGY 283
 QY 300 -----PSFKEMHAEVTVRVSQ-----OKAASLHE--LCAARVSEV----- 335
 Db 284 ILETNLQDLDPFLADKGFATTLQAAKPFSECRBAALAMQEBKGTTRKCSIGSIFT 343
 QY 336 --LQNRVHRTBEVKVDPAASYVDIAGVGLIDAERKGSILVGD--FEIAAKYVCTLE 391
 Db 344 PMLQGSFLATE-----NFFHNSKFFGLGEKEMLSMILAGKRFCEHWSKLVKYP 394
 QY 392 ETQPOSSPSCMDLNY--VSLLOEFGFP--RSKVLTUKTKI--LDNVTSMALGAT----- 441
 Db 395 TKDENLRYCPSAIIYISMLHDSGLVALDDEIRIKYASRAGEBDIPDMALGAFILNTAT 454
 QY 442 --FHYIDS 447
 Db 455 ATFDYSDN 462

Search completed: June 8, 2004, 09:50:35
 Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 09:48:14 ; Search time 18 Seconds
(without alignments)
1319.111 Million cell updates/sec

Title: US-09-905-589A-2

Perfect score: 2364
Sequence: 1 MKRISNHSRPAKVAIVPLG.....ALGAIIRHYIDSLNRQSPAS 456

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2364	100.0	484	1	ENP6_HUMAN
2	2063.5	87.3	455	1	ENP6_RAT
3	1007	42.6	427	1	ENP5_MOUSE
4	999	42.3	428	1	ENP5_HUMAN
5	959	40.6	469	1	ENP5_MESAU
6	498	21.1	518	1	GDAL1_YEAST
7	485	20.5	454	1	APV_SOLTU
8	442.5	18.7	455	1	NTPA_PEA
9	343	14.5	485	1	YTA4_CAEEL
10	329.5	13.9	630	1	YND1_YEAST
11	321.5	13.6	616	1	ENP4_HUMAN
12	320.5	13.6	613	1	ENP4_MOUSE
13	317.5	13.4	552	1	YBU4_CAEEL
14	315.5	13.3	493	1	ENP1_CHICK
15	314.5	12.7	529	1	ENP3_HUMAN
16	299.5	12.3	510	1	ENP1_MOUSE
17	294	12.4	495	1	ENP2_MOUSE
18	289	12.2	494	1	ENP2_CHICK
19	287	12.1	495	1	ENP2_RAT
20	286	12.1	511	1	ENP1_RAT
21	285.5	12.1	510	1	ENP1_HUMAN
22	265	11.2	495	1	ENP2_HUMAN
23	265	11.2	513	1	ENP1_BOVIN
24	253	10.7	510	1	ENP1_PIG
25	140.5	5.9	628	1	NTP2_TOXGO
26	139	5.9	628	1	NTP1_TOXGO
27	128.5	5.4	592	1	NTP4_TOXGO
28	109.5	4.6	505	1	MOO_ESERL
29	103.5	4.4	1516	1	CAIH_HUMAN
30	102.5	4.3	440	1	FTSA_BACSU
31	102	4.3	475	1	GATA1_THESA
32	102	4.3	623	1	DNAK_CAMGE
33	99.5	4.2	345	1	RUBB_PASMU

RESULT 1	ENP6_HUMAN	STANDARD	PRT	484 AA.
AC	075354	Q9UJD1		
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	ECtonucleoside triphosphate dihydrophosphate 6 (EC 3.6.1.6)			
DE	(NTPase6) (CD39 antigen-like 2).			
GN	ENTPD6 OR CD39L2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Keratinocytes;			
RX	MEDLINE=98341119; PubMed=9676430;			
RA	Chadwick B.P., Fritschau A.-M.;			
RT	"The CD39-like gene family: identification of three new human members (CD39L2, CD39L3, and CD39L4), their murine homologues, and a member of the gene family from Drosophila melanogaster.";			
RT	Genomics 50:357-367(1998).			
RU	[2]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=21638749; PubMed=11780052;			
RX	Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,			
RA	Jones M., Stavridis G., Almeida J.P., Babbage A.K., Baggaley C.L.,			
RA	Bailey J., Barlow K.P., Bates K.N., Beard L.M., Beare D.M.,			
RA	Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,			
RA	Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,			
RA	Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,			
RA	Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,			
RA	Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,			
RA	Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,			
RA	Graham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,			
RA	Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.V., Holden P.J.,			
RA	Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,			
RA	Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,			
RA	Lehvasaisto M.H., Leversha M.A., Lloyd D.M., Lovell J.D.,			
RA	Marshall V.L., Martin S.L., McConachie L.J., McKay K., McMurtry A.A.,			
RA	Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,			
RA	Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,			
RA	Phillimore B.J.C.T., Prichard J., Prichard S.R., Plumb R.W., Ramsay H.,			
RA	Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showken R., Sims S.,			
RA	Skuse C.D., Smith M.L., Soderlund C., Steward C.A., Suleston J.E.,			
RA	Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,			
RA	Tracey A., Tromans A.C., Vaudin R., Wall M., Wallis J.M.,			
RA	Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,			
RA	Willing L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,			
RA	Rogers U.;			
RT	"The DNA sequence and comparative analysis of human chromosome 20.";			
RT	Nature 414:865-871(2001).			
CC	-!- FUNCTION: Might support glycosylation reactions in the Golgi			
CC	apparatus and, when released from cells, might catalyze the			
CC	hydrolysis of extracellular nucleotides. Hydrolyses preferentially			

34	99.5	4.2	470	1	DDI1_BACST
35	99.5	4.2	2411	1	DAB_DROME
36	98.5	4.2	633	1	DNAX_AGRIS
37	98	4.1	415	1	SAHH_SULTO
38	97.5	4.1	415	1	SAHH_METTA
39	97.5	4.1	548	1	LIP2_CANRU
40	96	4.1	410	1	SAHH_THERO
41	96	4.1	711	1	TOP1_MYCEN
42	95.5	4.0	4074	1	PKHD_HUMAN
43	95	4.0	527	1	HSF8_LYCES
44	95	4.0	549	1	LIP1_CANRU
45	94.5	4.0	641	1	DNAX_XANNC

ALIGNMENTS

nucleoside 5'-diphosphates, nucleoside 5'-triphosphates are hydrolyzed only to a minor extent, there is no hydrolysis of nucleoside 5'-monophosphates. The order of activity with different substrates is GDP > IDP > UDP = CDP >> ADP (By similarity).

-/- CATALYTIC ACTIVITY: A nucleoside diphosphate + H(2)O = a nucleotide + phosphate.

-/- COFACTOR: Requires calcium and magnesium (By similarity).

-/- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI. BUT ALSO OCCURS IN A SOLUBLE EXTRACELLULAR FORM (BY SIMILARITY).

-/- TISSUE SPECIFICITY: Expressed in most tissues.

-/- SIMILARITY: Belongs to the GDAL / CD39 NTPase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF039916; AAC39883.1; -.
 EMBL; AL035252; CAB41571.1; -.
 Genew; HGNC:3368; ENTPD6.
 MIM; 603160; -.
 InterPro; IPR000407; GDAL_CD39_NTPase.
 Pfam; PF01150; GDAL_CD39_1.
 PROSITE; PS01238; GDAL_CD39_NTPASE; FALSE NEG.
 Hydrolase; Transmembrane; Glycoprotein; Calcium; Magnesium;
 Signal-anchor; Golgi stack.
 DOMAIN 1 39
 TRANSMEM 40 60
 CYTOPLASMIC (POTENTIAL).
 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 LUMENAL (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 220 220
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 284 284
 V -> L (IN REF. 2).
 CONFLICT 138 138
 CONFLICT 202 202
 E -> K (IN REF. 2).
 SEQUENCE 484 AA; 53233 MW; 27334B290DB8D64C CRC64;

Query Match 100.0%; Score 2364; DB 1; Length 484;
 Best Local Similarity 100.0%; Pred. No. 4e-180;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKISHGSLRVAKVAYPLGLCVGVFIYVAYIKWHRATQATQAFSTITRAAPGARWQQA 60
 DB 29 MRKISHGSLRVAKVAYPLGLCVGVFIYVAYIKWHRATQATQAFSTITRAAPGARWQQA 88
 QY 61 SPLGTAADGHEVYFGIMFDAGSTGRVHVQFTRPREPTTLTHETFKAVKPGLSAYADD 120
 DB 89 SPLGTAADGHEVYFGIMFDAGSTGRVHVQFTRPREPTTLTHETFKAVKPGLSAYADD 148
 QY 121 VEKSAQGIREDLDVAKQDIPDFWFKATPLVLKATAGLLPGEKAQKLLQKVKEVFKASP 180
 DB 149 VEKSAQGIREDLDVAKQDIPDFWFKATPLVLKATAGLLPGEKAQKLLQKVKEVFKASP 208
 QY 181 FLVGGDCVSIKNGTDEGVSAWITINFLGSLKTPGGSSVGMLDLGGGQTQAFPRVEGT 240
 DB 209 FLVGGDCVSIKNGTDEGVSAWITINFLGSLKTPGGSSVGMLDLGGGQTQAFPRVEGT 268
 QY 241 LQASPPGYLTALRMFNRYKLSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 300
 DB 269 LQASPPGYLTALRMFNRYKLSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 328
 QY 301 SPKGEWEHAEVYRVSQKAAASLHELCAARVSEVLQNRVHRTBEVKHVDYAFSYTYDL 360
 DB 329 SPKGEWEHAEVYRVSQKAAASLHELCAARVSEVLQNRVHRTBEVKHVDYAFSYTYDL 388
 QY 361 AAGVGLIDAEKGGSLVGDGFETAAKVCRTLETQPSQSPFSCMDLTYYVSLLLQFGFPFRS 420
 DB 389 AAGVGLIDAEKGGSLVGDGFETAAKVCRTLETQPSQSPFSCMDLTYYVSLLLQFGFPFRS 448
 QY 421 KVLKLRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456

Db 449 KVLKLRKIDNVETSWALGAIFHYIDSLNRQKSPAS 484

RESULT 2

ENP6_RAT STANDARD; PRT; 455 AA.

ID ENP6_RAT
 AC Q9BR31;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DB Ectonucleoside triphosphate diphosphohydrolase 6 (EC 3.6.1.6)
 DE (NTPDase6) (CD39 antigen-like 2).
 GN ENTPD6 OR CD39L2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RX NCBI_TaxID=10116;
 RX NCBI_TaxID=10116;
 RP SEQUENCE FROM N.A. SUBCELLULAR LOCATION, AND CHARACTERIZATION.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=20498744; PubMed=11042118;
 RA Braun N., Fengler S., Ebeling C., Servos J., Zimmermann H.;
 RT "Sequencing, functional expression and characterization of rat
 NTPDase6, a nucleoside diphosphatase and novel member of the ecto-
 nucleoside triphosphate diphosphohydrolase family."; Biochem. J. 351:639-647(2000).
 RL Biochem. J. 351:639-647(2000).
 CC -/- FUNCTION: Might support glycosylation reactions in the Golgi apparatus and, when released from cells, might catalyze the hydrolysis of extracellular nucleosides. Hydrolyzes preferentially nucleoside 5'-diphosphates, nucleoside 5'-triphosphates are hydrolyzed only to a minor extent, there is no hydrolysis of nucleoside 5'-monophosphates. The order of activity with different substrates is GDP > IDP >> UDP = CDP >> ADP.
 CC -/- CATALYTIC ACTIVITY: A nucleoside diphosphate + H(2)O = a nucleotide + phosphate.
 CC -/- COFACTOR: Requires calcium and magnesium.
 CC -/- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI. BUT ALSO OCCURS IN A SOLUBLE EXTRACELLULAR FORM.
 CC -/- SIMILARITY: Belongs to the GDAL / CD39 NTPase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AJ277748; CAC16598.1; -.
 InterPro; IPR000407; GDAL_CD39_NTPase.
 Pfam; PF01150; GDAL_CD39_1.
 PROSITE; PS01238; GDAL_CD39_NTPASE; FALSE NEG.
 Hydrolase; Transmembrane; Glycoprotein; Calcium; Magnesium;
 Signal-anchor; Golgi stack.
 DOMAIN 1 12
 TRANSMEM 13 32
 CYTOPLASMIC (POTENTIAL).
 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 LUMENAL (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 192 192
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 256 256
 SEQUENCE 455 AA; 49899 MW; 19A22E8BAEF0F77B CRC64;

Query Match 87.3%; Score 2063.5; DB 1; Length 455;
 Best Local Similarity 86.6%; Pred. No. 2.7e-156;
 Matches 394; Conservative 27; Mismatches 33; Indels 1; Gaps 1;

QY 1 MRKISHGSLRVAKVAYPLGLCVGVFIYVAYIKWHRATQATQAFSTITRAAPGARWQQA 60
 DB 1 MRKISHGSLRVAKVAYPLGLCVGVFIYVAYIKWHRATQATQAFSTITRAAPGARWQQA 60
 QY 61 SPLGTAADGHEVYFGIMFDAGSTGRVHVQFTRPREPTTLTHETFKAVKPGLSAYADD 120
 DB 61 SPLGTAADGHEVYFGIMFDAGSTGRVHVQFTRPREPTTLTHETFKAVKPGLSAYADD 120

QY	12	VESSAGIGIEELIDVAADQ:PPDFPMKATPLVYKATAGLRLTGEKAAKLLQKVEVFYAS	180
Db	121	VESSAGIGIEELINVAQKHPIYDFMKATPLVYKATAGLRLTGEKAAKLLQKVEVFYAS	180
QY	181	FLVGGDCVSIANGTDEGVSAWITTFNLFGSLKTPGGSSVGMULDGGSTQIAFLPREGT	240
Db	181	FLVGGDCVSIANGTDEGVSAWITTFNLFGSLKTPGGSSVGMULDGGSTQIAFLPREGT	240
QY	241	LQASPFGVLTALMPFRITKLSYSYELGLGLMSALALTLGVEGQAPADGKEIVSPCLSP	300
Db	241	LQASPFGVLTALMPFRITKLSYSYELGLGLMSALALTLGVEGQAPADGKEIVSPCLSP	300
QY	301	SPFGEMHEAEVTVYRVSQGRPAASLHELCAARVSEVLQNRVHRTVEYKVDVFAFSYYDI	360
Db	301	SPFGEMHEAEVTVYRVSQGRPAASLHELCAARVSEVLQNRVHRTVEYKVDVFAFSYYDI	360
QY	361	AAAGVGLIDAEKGSILVDPFELAAKYRCVLTETQSSPFGOMLTIVSYLLDGFPPRS	420
Db	361	AAAGVGLIDAEKGSILVDPFELAAKYRCVLTETQSSPFGOMLTIVSYLLDGFPPRS	420
QY	421	KVFKLTRKIDNVEVTSWALGAFHYIDSLNRQKSPA	455
Db	421	KVFKLTRKIDNVEVTSWALGAFHYIDSLNRQKSPA	455
RESULT 3			
ENPS_MOUSE		STANDARD;	PRT; 427 AA.
ID	ENPS_MOUSE		
AC	Q9WU29; O70214;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	ECconuclease di triphosphate di phosphonucleotidase 5 precursor		
DE	(EC 3.6.1.6) (NTPases) (Nucleoside diphosphatase) (CD39 antigen-like		
DE	4) (ER-UDPase)."		
GN	ENTPDS OR CD39L4.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Testis;		
RC	MEDLINE=98119025; PubMed=9457681;		
RA	Chadwick B.P., Williamson J., Sheer D., Frieschaut A.M.;		
RT	"CDNA cloning and chromosomal mapping of a mouse gene with homology to		
RT	NTPases.";		
RL	Mamm. Genome 9:162-164(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND CHARACTERIZATION.		
RC	TISSUE=Liver;		
RC	MEDLINE=99298181; PubMed=10369669;		
RA	Tomobetta E.S., Helenius A.;		
RT	"Glycoprotein regucosylation and nucleotide sugar utilization in the		
RT	secretory pathway: identification of a nucleoside diphosphatase in the		
RT	endoplasmic reticulum.";		
RL	EMBO J. 18:3282-3292(1999).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Kidney;		
RC	MEDLINE=1085560; PubMed=11117651;		
RA	Kawai J., Shinagawa A., Shitara K., Yoshino M., Itoh M., Ishii Y.,		
RA	Arakawa T., Hara A., Fukunishi Y., Komno H., Adachi J., Fukuda S.,		
RA	Aitawa K., Izawa Y., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,		
RA	Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Gissi C., Kang B., Kochwa H.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,		
RA	Schirral L.M., Staudli F., Suzuki R., Tomita M., Wagner I., Maehino T.,		
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bazer G.,		
RA	Blake U., Botelli D., Bojungna N., Cammici P., de Bonardo M.F.,		
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,		

RA Gustinich S., Hill D., Hofmann M., Hune D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombasche P., Nordepe P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toy-o-ka K., Wang K.H., Wetz C., Whitaker C., Wilting L., Wyshawski S., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S., Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection." ;
RN Nature 409:665-690(2001).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altshuler S.F., Zeeberg B., Bluetworth K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Datchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uebli T.B., Toshiyuki S., Carninci P., Prange C., Rana S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guarante P.H., Richards S., Wooley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs B.A., Winkler D., Heltzer J., Heltzer E., Kettner M., Mader A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shenyuk G., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Rutterfield Y.S.N., Krzyzanski M.I., Skalska U., Small D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences." ;
RL Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).
CC -1- FUNCTION: Likely to promote reglycosylation reactions involved in glycoproteins folding and quality control in the endoplasmic reticulum. Hydrolyzes UDP- GDP AND IDP but not any other nucleoside di-, mono or triphosphates, nor thiamine pyrophosphate.
CC -1- CATALYTIC ACTIVITY: A nucleoside diphosphate + H(2)O = a nucleotide + phosphate.
CC -1- COFACTOR: Requires calcium and magnesium.
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- PTM: Glycosylated with high mannose N-linked glycans.
CC -1- MISCELLANEOUS: Optimal pH is neutral.
CC -1- SIMILARITY: Belongs to the GDA1 / CD39 NTPase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcement/> or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AF0066482; AAC05181.1; -
DR EMBL; AJ238636; CAB45533.1; -
DR EMBL; AK002618; BA822234.1; -
DR EMBL; BC015247; AA815247.1; -
DR MGd; MG1:1321385; Eucp45.
DR InterPro; IPRO00407; GDA1_CDS9_NTPase.
DR Pfam; PF01150; GDA1_CDS9_1.
DR PROSITE; PS01238; GDA1_CDS9_NTPase; FALSE NEG.
KM Hydroxylase; Transmembrane; Glycoprotein; Calcium; Magnesium; Endoplasmic reticulum; Signal.
FT CHAIN 1 18
FT SIGNAL 1 18
FT CHAIN 19 427
FT SIGNAL 19 427
FT CHAIN 20 427
FT SIGNAL 20 427
FT CHAIN 21 427
FT SIGNAL 21 427
FT CHAIN 22 427
FT SIGNAL 22 427
FT CHAIN 23 427
FT SIGNAL 23 427
FT CHAIN 24 427
FT SIGNAL 24 427
FT CHAIN 25 427
FT SIGNAL 25 427
FT CHAIN 26 427
FT SIGNAL 26 427
FT CHAIN 27 427
FT SIGNAL 27 427
FT CHAIN 28 427
FT SIGNAL 28 427
FT CHAIN 29 427
FT SIGNAL 29 427
FT CHAIN 30 427
FT SIGNAL 30 427
FT CHAIN 31 427
FT SIGNAL 31 427
FT CHAIN 32 427
FT SIGNAL 32 427
FT CHAIN 33 427
FT SIGNAL 33 427
FT CHAIN 34 427
FT SIGNAL 34 427
FT CHAIN 35 427
FT SIGNAL 35 427
FT CHAIN 36 427
FT SIGNAL 36 427
FT CHAIN 37 427
FT SIGNAL 37 427
FT CHAIN 38 427
FT SIGNAL 38 427
FT CHAIN 39 427
FT SIGNAL 39 427
FT CHAIN 40 427
FT SIGNAL 40 427
FT CHAIN 41 427
FT SIGNAL 41 427
FT CHAIN 42 427
FT SIGNAL 42 427
FT CHAIN 43 427
FT SIGNAL 43 427
FT CHAIN 44 427
FT SIGNAL 44 427
FT CHAIN 45 427
FT SIGNAL 45 427
FT CHAIN 46 427
FT SIGNAL 46 427
FT CHAIN 47 427
FT SIGNAL 47 427
FT CHAIN 48 427
FT SIGNAL 48 427
FT CHAIN 49 427
FT SIGNAL 49 427
FT CHAIN 50 427
FT SIGNAL 50 427
FT CHAIN 51 427
FT SIGNAL 51 427
FT CHAIN 52 427
FT SIGNAL 52 427
FT CHAIN 53 427
FT SIGNAL 53 427
FT CHAIN 54 427
FT SIGNAL 54 427
FT CHAIN 55 427
FT SIGNAL 55 427
FT CHAIN 56 427
FT SIGNAL 56 427
FT CHAIN 57 427
FT SIGNAL 57 427
FT CHAIN 58 427
FT SIGNAL 58 427
FT CHAIN 59 427
FT SIGNAL 59 427
FT CHAIN 60 427
FT SIGNAL 60 427
FT CHAIN 61 427
FT SIGNAL 61 427
FT CHAIN 62 427
FT SIGNAL 62 427
FT CHAIN 63 427
FT SIGNAL 63 427
FT CHAIN 64 427
FT SIGNAL 64 427
FT CHAIN 65 427
FT SIGNAL 65 427
FT CHAIN 66 427
FT SIGNAL 66 427
FT CHAIN 67 427
FT SIGNAL 67 427
FT CHAIN 68 427
FT SIGNAL 68 427
FT CHAIN 69 427
FT SIGNAL 69 427
FT CHAIN 70 427
FT SIGNAL 70 427
FT CHAIN 71 427
FT SIGNAL 71 427
FT CHAIN 72 427
FT SIGNAL 72 427
FT CHAIN 73 427
FT SIGNAL 73 427
FT CHAIN 74 427
FT SIGNAL 74 427
FT CHAIN 75 427
FT SIGNAL 75 427
FT CHAIN 76 427
FT SIGNAL 76 427
FT CHAIN 77 427
FT SIGNAL 77 427
FT CHAIN 78 427
FT SIGNAL 78 427
FT CHAIN 79 427
FT SIGNAL 79 427
FT CHAIN 80 427
FT SIGNAL 80 427
FT CHAIN 81 427
FT SIGNAL 81 427
FT CHAIN 82 427
FT SIGNAL 82 427
FT CHAIN 83 427
FT SIGNAL 83 427
FT CHAIN 84 427
FT SIGNAL 84 427
FT CHAIN 85 427
FT SIGNAL 85 427
FT CHAIN 86 427
FT SIGNAL 86 427
FT CHAIN 87 427
FT SIGNAL 87 427
FT CHAIN 88 427
FT SIGNAL 88 427
FT CHAIN 89 427
FT SIGNAL 89 427
FT CHAIN 90 427
FT SIGNAL 90 427
FT CHAIN 91 427
FT SIGNAL 91 427
FT CHAIN 92 427
FT SIGNAL 92 427
FT CHAIN 93 427
FT SIGNAL 93 427
FT CHAIN 94 427
FT SIGNAL 94 427
FT CHAIN 95 427
FT SIGNAL 95 427
FT CHAIN 96 427
FT SIGNAL 96 427
FT CHAIN 97 427
FT SIGNAL 97 427
FT CHAIN 98 427
FT SIGNAL 98 427
FT CHAIN 99 427
FT SIGNAL 99 427
FT CHAIN 100 427
FT SIGNAL 100 427
FT CHAIN 101 427
FT SIGNAL 101 427
FT CHAIN 102 427
FT SIGNAL 102 427
FT CHAIN 103 427
FT SIGNAL 103 427
FT CHAIN 104 427
FT SIGNAL 104 427
FT CHAIN 105 427
FT SIGNAL 105 427
FT CHAIN 106 427
FT SIGNAL 106 427
FT CHAIN 107 427
FT SIGNAL 107 427
FT CHAIN 108 427
FT SIGNAL 108 427
FT CHAIN 109 427
FT SIGNAL 109 427
FT CHAIN 110 427
FT SIGNAL 110 427
FT CHAIN 111 427
FT SIGNAL 111 427
FT CHAIN 112 427
FT SIGNAL 112 427
FT CHAIN 113 427
FT SIGNAL 113 427
FT CHAIN 114 427
FT SIGNAL 114 427
FT CHAIN 115 427
FT SIGNAL 115 427
FT CHAIN 116 427
FT SIGNAL 116 427
FT CHAIN 117 427
FT SIGNAL 117 427
FT CHAIN 118 427
FT SIGNAL 118 427
FT CHAIN 119 427
FT SIGNAL 119 427
FT CHAIN 120 427
FT SIGNAL 120 427
FT CHAIN 121 427
FT SIGNAL 121 427
FT CHAIN 122 427
FT SIGNAL 122 427
FT CHAIN 123 427
FT SIGNAL 123 427
FT CHAIN 124 427
FT SIGNAL 124 427
FT CHAIN 125 427
FT SIGNAL 125 427
FT CHAIN 126 427
FT SIGNAL 126 427
FT CHAIN 127 427
FT SIGNAL 127 427
FT CHAIN 128 427
FT SIGNAL 128 427
FT CHAIN 129 427
FT SIGNAL 129 427
FT CHAIN 130 427
FT SIGNAL 130 427
FT CHAIN 131 427
FT SIGNAL 131 427
FT CHAIN 132 427
FT SIGNAL 132 427
FT CHAIN 133 427
FT SIGNAL 133 427
FT CHAIN 134 427
FT SIGNAL 134 427
FT CHAIN 135 427
FT SIGNAL 135 427


```

FT TRANSMEM 10 24 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 25 518 LUMENAL (POTENTIAL).
FT CARBOHYD 41 41 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 335 335 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 518 AA; 56821 MW; 9A61182D5ED2AD2ADF CRC64;

Query Match 21.1%; Score 498; DB 1; Length 518;
Best Local Similarity 33.1%; Pred. No. 7.1e-32;
Matches 144; Conservative 60; Mismatches 153; Indels 78; Gaps 15;

QY 65 TRADGHEVYGLMFDAGSTGTRVHVQF---TRPRETETLTHETFKAVKPGLSAYADV 121
DB 86 TCSEEHK--YVIMIDAGSTGSRVHYKFDVCHSP-----PTLDEKFDMLPEGLSFDYD 139
QY 122 EKSAQGIKRELLDVAKODIPDFFWKATPLVLKATAGLRLLPGEKAQKLOKVE--YFKASP 180
DB 140 VGAANSLDPLLVAMNYVPIKARSTPVAVKATAGLRLLGDAKSKILSAVRDHLEKDP 199
QY 181 F-LVGDDCVSINMGTDGVSAMTINFLGSLKTPGGS--SVGMLDLGGSTQIAFLPRV 237
DB 200 FPVEGDSVIMGDEGVFAMITNLYLNGNIGANGPKLPTAAVFDLGGGSGQIVFEPTF 259
QY 238 EGTQASPGYLTALMFMNRTKLYSYGLGLMSARLAILGGVGEOPAKDGK----- 291
DB 260 PINEKMDVGEKFDLFGDENTLYQFSLGYGLKEGRKNVSNLVENALKDGKILKGN 319
QY 292 -----ELVSPCLSPSFKGEW-----HAETVTV-----SQKAAASLHELCAARVSEV 335
DB 320 TKNHQLSSPCLPKVKNATNEKVTLESKTYTDFIGDPSPGAQ-----CRFLTDEI 371
QY 336 LQ-----NRVHRTTEEKVV-----DFYAFSYVYDLAGVGLIDAEGKGLWVGD 379
DB 372 LNKDAQCCQSPPCFNGVHGPSLVRTKESNDIYFSYFDRTRPLQM-----PLSFTLINE 426
QY 380 FETAAYVKTLET-----QPQSPFSCMDLTY-VSLLQEFPGPRSKVLK 425
DB 427 LNDLARIYCKGEETWNSVFSRAGSGLDELSHFLCLDSFQVSLIHTGYDIPLQELRT 486
QY 426 TRKIDNVEITSWALGA 440
DB 487 GKIANKEIGWCLGA 501

RESULT 7
APY_SOLTU STANDARD; PRT; 454 AA.
AC P80536; Q43164;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Apyrase precursor (EC 3.6.1.5) (ATP-diphosphatase) (Adenosine
DE diphosphatase) (ADPase) (ATP-diphosphohydrolase).
GN R80P1.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 59-160; 236-253 AND 332-345.
RC TISSUE= tuber;
RX MEDLINE=96158985; PubMed=8579614;
RA Handa M., Guidotti G.;
RT "Purification and cloning of a soluble ATP-diphosphohydrolase
RT (aprase) from potato tubers (Solanum tuberosum).";
RL Biochem. Biophys. Res. Commun. 218:916-923(1996).
RN [2]
RP SEQUENCE OF 42-54; 68-95 AND 236-253.
RC STRAIN=cv. Desiree;
RX MEDLINE=96355615; PubMed=8703025;
RA Vasconcelos E.G., Ferreira S.T., de Carvalho T.M.U., de Souza W.,

```

```

RA Kettlun A.M., Mancilla M., Valenzuela M.A., Verjovski-Almeida S.;
RA "Partial purification and immunohistochemical localization of ATP
RA diphosphohydrolase from Schistosoma mansoni. Immunological cross-
RA reactivities with potato apyrase and Toxoplasma gondii nucleoside
RA triphosphate hydrolase.";
RA J. Biol. Chem. 271:22139-22145(1996).
CC -!- FUNCTION: Catalyzes the hydrolysis of phosphoanhydride bonds of
CC nucleoside tri- and di-phosphates.
CC -!- CATALYTIC ACTIVITY: ATP + 2 H(2)O = AMP + 2 phosphate.
CC -!- COFACTOR: Calcium.
CC -!- SUBCELLULAR LOCATION: Membrane-associated (Probable).
CC -!- SIMILARITY: Belongs to the GDAL / CD39 NTPase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U58597; AAB02720.1; -.
CC PIR; JC4616; JC4616.
CC InterPro; IPR000407; GDAL_CD39_NTPase.
CC Pfam; PF01150; GDAL_CD39; 1.
CC PROSITE; PS01238; GDAL_CD39_NTPASE; 1.
KW Hydrolase; Transmembrane; Calcium; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 454 APYRASE.
FT TRANSMEM 426 446 POTENTIAL.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 454 AA; 50041 MW; 9D9EFE431DA2F52F CRC64;

Query Match 20.5%; Score 485; DB 1; Length 454;
Best Local Similarity 31.9%; Pred. No. 6.4e-31;
Matches 138; Conservative 68; Mismatches 183; Indels 44; Gaps 12;

QY 59 AHSPLGTAAADGHE--VFYGMFDAGSTGTRVHVQFTRPPRETPTLTH-ETFKAVKPGLSA 116
DB 30 AQPLRRLHLLSHSEHYAVIFDAGSTGSRVHVFRFDEKGLLLPIGNNIEYFMATEPGLSS 89
QY 117 YADDEKSAQGIKRELLDVAKODIPDFFWKATPLVLKATAGLRLLPGEKAQKLOKVEVF 176
DB 90 YADPKAANSLPFLLDGAEVVPQELQSETPLGATAGLRMLKGDAEKILQAVRNLV 149
QY 177 K-ASPELVGDDCVSINMGTDGVSAMTINFLGSLKTPGGSVGMGLDGGSTQIAFLP 235
DB 150 KNGTTFHSDQWVITLDTGTOEGSYMMAINLYLGNLQKYKSTIATIDLGGSVQWYAI 209
QY 236 RVETGLQA-----SPPGYLTALRMFNRTKLYSYGLGLMSARLAILGGVGEOPAKDGK 291
DB 210 SNEQFAKAPQNEDEGPYVQCKHLMKSDYNLYVHSYLYNGOLAGRAEIF-----KASR 261
QY 292 ELVSPCLSPSFKGEWHAETVTVYVSGKAAASLHELCAARVSEVLQ-NRVHRTTEEKVV--- 347
DB 262 NESNPCALEGCDGYISYGGVDYKVPKPKGSS-WKCCRRLTRHALKINAKNTEECTFNG 320
QY 348 -----HVDYFARSYVYDLAGVGLIDAEGKGLWVG-DFEIAAKYVCR----- 390
DB 321 VNWGGGDDGQKNTHASSFFDYIGAQVIVDTKFPSSALAKPIQVLYNAKAVCAQTNVADIKS 380
QY 391 --LETQSSQSPFSCMDLTY-VSLLQEFPGPRSKVLKTRKIDN----VETSWALGAIFH 443
DB 381 IFPKTDQRNIPYLCMDLIYETLLVDGFGNLNPKHEITVHDVQKNYLVGAAPLGAID 440
QY 444 VIDSLNPKQSPAS 456
DB 441 LVSSTTKIRVAS 453

RESULT 8

```

NTPA_PEA
ID NTPA_PEA STANDARD; PRT; 455 AA.
AC P52914;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nucleoside-triphosphatase (EC 3.6.1.15) (Nucleoside triphosphate
phosphohydrolase) [NTPase] (Apyrase).
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxId=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Alaska; TISSUE=Plumule;
RX MEDLINE=96197404; Pubmed=8616230;
RA Heide H., Tong C.G., Thomas C., Roux S.J.;
RT "Light-modulated abundance of an mRNA encoding a calmodulin-regulated,
RT chromatin-associated NTPase in pea."
RT Plant Mol. Biol. 30:135-147(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Alaska; TISSUE=Stem;
RA Shibata K., Abe S., Davies E.;
RT "Structure of the coding region and mRNA variants of the apyrase from
RT Pisum sativum."
RT Acta Physiol. Plant. 20:3-13(2001).
CC -1- FUNCTION: Might be involved in RNA transport out of nuclei.
CC -1- CATALYTIC ACTIVITY: NTP + H(2)O = NDP + phosphate.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the GDA1 / CD39 NTPase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z32743; CAAB3655.1; -
DR EMBL; AB027614; BAAB8900.1; -
DR PIR; S65147; S46859.
DR InterPro; IPR000407; GDA1_CD39_NTPase.
DR Pfam; PF01150; GDA1_CD39; 1.
DR PROSITE; PS01238; GDA1_CD39_NTPASE; 1.
KW Hydrolyase; Nuclear protein.
SQ SEQUENCE 455 AA; 50072 MW; 50FDF0023AB04299 CRC64;

Query Match 18.7%; Score 442.5; DB 1; Length 455;
Best Local Similarity 29.2%; Pred. No. 1.5e-27;
Matches 125; Conservative 79; Mismatches 145; Indels 79; Gaps 15;

QY 74 YGIMEDAGSTGTRGVHVFQFRPPRPPTLTH-----ETFAVVRGSLGSAVADVEKSAOGI 128
DB 44 YAVVFDAGSTGSRKHVHVF-----NQNDLHLHGKGVYVYKTPGSLSYANNEQAAKSL 99
QY 129 RELIDVAKODIPDFFWKATPVLVKATAGLRLPGEKAKLLQKVEVF-KASPTLVGDDC 187
DB 100 IPLLQAEADVDPDLQCPTRVRLGATAGLRLNLDGASEKILQSRDMLSNRSTFNQPD 159
QY 168 VSINMGDEGVASMTINFLGSLKTPGSSVGMGLDGGSTQIAF-----LPRVE 238
DB 160 VSIIIDGDEGVSLVTVYVYVAGNLGKRYTKVGTIDLGSSVQVAIVASKTKAKAPKV- 218
QY 239 GTLQASPGYLTALRMFRRTYKLYSYGLGLMSARLALIGVGGPAPADGKELVSPCL 298
DB 219 --ADGDDP-VTKVVLKGIPIYDLVYHSLHFGREASAEILKLTTPRP-----NPCL 267
QY 299 SPFSKGEHEHAEVTVYRVSQGAASLHELCAARVSEVLQNRVHRTVEVKA----- 348
DB 268 LAGNGIYTVYSGEERKATVATYSGANFNK-----CKQITRKALNTPCPYQNTCTFG 318

QY 349 -----VDFYASYYDLAAGVGLIDAEKGSLLV-VGPEIAKAYVCR-TLETOP 395
DB 319 GINWGGGNGQKNLFASSSFYLPEDTGVWDASTPNTLKPVDLETAKERACALNFEDAK 378
QY 396 QSSPF-----SCMDLTFVSLILOEFGFPRSKVLKT--RKIDN-----VETSMALGA 440
DB 379 STYFFLQKXNVASVYCWDLIYQVLLVD-GFGDLP.LKITSKEIERQDAIVEAWPLGN 437
QY 441 IFHTYIDL 448
DB 438 AVEAIAL 445

RESULT 9
Y44E_CABEL
ID Y44E_CABEL STANDARD; PRT; 485 AA.
AC Q18411;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 54.3 kDa protein C33H5.14 in chromosome IV.
DE C33H5.14.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Bradshaw H., Steillyes L.;
RT Submitted (Dec-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the GDA1 / CD39 NTPase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U41007; AAB82272.1; -
DR PIR; T34147; T34147.
DR WormPep; C33H5.14; CE04157.
DR InterPro; IPR000407; GDA1_CD39_NTPase.
DR Pfam; PF01150; GDA1_CD39; 1.
DR PROSITE; PS01238; GDA1_CD39_NTPASE; 1.
KW Hypothetical protein; Transmembrane; Hydrolyase.
FT TRANSMEM 439 AA; POTENTIAL.
SQ SEQUENCE 485 AA; 54309 MW; D864D1ADC20F5815 CRC64;

Query Match 14.5%; Score 343; DB 1; Length 485;
Best Local Similarity 25.8%; Pred. No. 1.3e-19;
Matches 113; Conservative 74; Mismatches 157; Indels 94; Gaps 16;

QY 70 HEVFGIMEDAGSTGTRGVHVFQFRPP-----REPTPLTHET--FKAVKGLSAVADV 121
DB 21 NNIRKGVICDAGSSSTR--LFFVYTLKPSGGLTNIDTLHSEBPVVKVTDGLSPFGKP 78
QY 122 EKSAGGIELIDVAKODIPDFFWKATPVLVKATAGLRLPGEKAKLLQKVEVFKAAPF 181
DB 79 EGVEVYLP.LRFASBEH.PYEGIGETDILLIFATGMRLLPFAQDAIYKLNQNGKSVTA 138
QY 182 L-VGDDCVISINMGDEGVASMTINFLGSLKTPGSSVGMGLDGGSTQIAFLPRVGT 240
DB 139 LRVSSNRIIDGAGEGVSYTAIVYILGRPDKENDSKYGMIDWGASVQIAF--ELANE 196
QY 241 LQASPGYLTALRM-----FNRTYKLYSYGLGLMSARLALIGVGGPAPADGKELV 294
DB 197 KESYNGGVNVEINLQSIETNEDYKTKISTTFILGAGAG--LKKYENLVYSGNS-- 250
QY 295 SPCLSP-----SFKGEHEHAEVTVYRVSQGAASLHELCAARVSEV----- 335

```

Db 251 NUSCSPRGLNRLIGETVNTGTGEW-----DVCLAQVSLIGDKAQP 291
QY 336 -----LQNRVHRTTEEVKHYDFAFSYYYDLAAGVGLIDAEGKGLVVGDFEIAAKYV 387
Db 292 SCNPPTCFLRNVIAPSVNLSTVQLYGFSEYWTTSNFG-----SGEYHYQKFTDEVKRY 346
QY 388 CR-----TLTQOSPPF-----SCMDTYVSLILOEFGPFRSKVLKTR---KID 430
Db 347 CQKMDWDIQGFKRNEFPFNADIERLGTNCFKAAWTVSLHD-GFNVDKTKHLFQSVLKA 405
QY 431 NVETSWALGAIFHYIDSL 448
Db 406 GEEMQWALGNMLYHSKDL 423

RESULT 10
YND1_YEAST
ID YND1_YEAST STANDARD; PRT; 630 AA.
AC P40009;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Golgi apyrase (EC 3.6.1.5) (ATP-diphosphatase) (Adenosine
DE diphosphatase) (ADPase) (ATP-diphosphohydrolase) (Golgi nucleoside
DE diphosphatase).
GN YND1 OR YERO05W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=S288C;
RX MEDLINE=99340091; PubMed=10409709;
RA Gao X.D., Kaigorodov V., Jigami Y.;
RT "YND1, a homologue of GDAL, encodes membrane-bound apyrase required
RT for Golgi N- and O-glycosylation in Saccharomyces cerevisiae.";
RL J. Biol. Chem. 274:21450-21456(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313264; PubMed=9169868;
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
RA Hunick-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA Lin D., Moesdale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RL Nature 387:78-81(1997).
CC -I- FUNCTION: Catalyzes the hydrolysis of phosphoanhydride bonds of
CC nucleoside tri- and di-phosphates. Has equal high activity toward
CC ADP/ATP, GMP/GTP, and UTP/UTP and approximately 50% less toward
CC CDP/CTP and thianine pyrophosphate. Has no activity toward GMP.
CC Required for Golgi glycosylation and cell wall integrity.
CC -I- CATALYTIC ACTIVITY: ATP + 2 H(2)O = AMP + 2 phosphate.
CC -I- PATHWAY: Glycosylation.
CC -I- SUBCELLULAR LOCATION: Golgi; membrane-bound.
CC -I- SIMILARITY: Belongs to the GDAL / CD39 NTPase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; AP203695; AAF17573.1; -.
CC ENBL; U18776; AAB64538.1; -.
CC PIR; S50463; S50463.

```

```

DR Germline; 139085; -.
DR SGD; S000807; YND1.
DR GO; GO:0017110; F: Nucleoside diphosphatase activity; IDA.
DR GO; GO:0006486; P: Protein amino acid glycosylation; IMP.
DR InterPro; IPR000407; GDAL_CD39_NTPase.
DR Pfam; PF01150; GDAL_CD39_1.
DR PROSITE; PS01238; GDAL_CD39_NTPase; 1.
KW Hydrolase; Transmembrane; Golgi stack.
FT DOMAIN 1 500 LUMENAL (POTENTIAL).
FT TRANSMEM 501 517 POTENTIAL.
FT DOMAIN 518 630 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 630 AA; 71851 MW; 02F8D24A78212544 CRC64;

Query Match 13.9%; Score 329.5; DB 1; Length 630;
Best Local Similarity 26.9%; Pred. No. 2.e-18;
Matches 122; Conservative 72; Mismatches 163; Indels 97; Gaps 22;

QY 74 YGIMFDAGSTGTRVHVVFQFTRPPRTPTLTHE-----TFKAVKP 112
Db 10 FGIVIDAGSSGSIHVFKW---QDTESLLHATNQDSQILQSVPHIQEKDWTFF-LNP 64
QY 113 GLSAYADDVEKSAQG-----IRELLVAKODIPDFEWKATPLVLKATAGLLPGEKAQK 167
Db 65 GLSSF-----EKKPQDAYKSHIKPLDFAKNIIPESHWSSCFVIQATAGMRLLPQDIQSS 120
QY 168 LLOKVKEVFK-ASPFVLVGDDC---VSIMMGTDREGSAWITINFLTGLSK--TPGGS---S 218
Db 121 ILDGLCQLGKHPAEFLV-EDCSAQIQVIDGETEGYGLNLYGHENDYNEVSDHPT 179
QY 219 VGMLDLGGSTQIAFLPRVEGTLOAQPFGVLTG-LRMFN---RTYKLYSYSLGLGMSA 274
Db 180 FGFMDMGASTQIAFAPHDSGEIARHDDIATIFLSVNGDLQKWDVVFVSTLWGFQANQA 239
QY 275 R---LAILGGVEGQAKD-----GKELVSPCLSPSKFGKEWHAETVTVRSVSG---QKA 320
Db 240 RRYLAQLINTLPENTDNDYENDDFSTNLDNDCMPRGSSDTDFEKTIFHIAGSGNYEQC 299
QY 321 AASHELCAARV---SEVLQNRVHRT-EVKHVDYAFSYYYDLAAGVGLIDAEGKGL 375
Db 300 TKSITYPILLKNMPCDDDEPCLFNGVHAPRIDFANDKFGTSEYWTYANDV---FKLGEY 355
QY 376 VVGDFEIAAKYVCRTLETQ-----PQS-SPFSCMDLTVYVSLLLQD-FGFPRS 420
Db 356 NDFKFSKSLREFCNSNWTQLANSKGVNISPENFLKDACFKGNVWLILHGEFDMPRI 415
QY 421 KV-----LKLTRKIDNVTSWALGAIPHY 444
Db 416 DVDANVNDRLPQSVKEVEERELSLWTIGRILLY 449

RESULT 11
ENP4_HUMAN
ID ENP4_HUMAN STANDARD; PRT; 616 AA.
AC Q9Y27; O15092;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 4 (EC 3.6.1.6)
DE (NTPDase4) (Uridine-diphosphatase) (UDPase) (Lysosomal apyrase-like
DE protein of 70 kDa).
GN LYSALI OR ENTPD4 OR LALP70 OR KIAA0392.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=98225229; PubMed=95566635;
RA Wang T.-P., Guidotti G.;
RT "Golgi localization and functional expression of human uridine
RT diphosphatase.";
RL J. Biol. Chem. 273:11392-11399(1998).

```

RE SEQUENCE FROM N.A. (ISOFORM 1).
 RC TTSUS-Pancreatic adenocarcinoma;
 RX MEDLINE=99329914; PubMed=10399803;
 RA Biededick A., Rose S., Elsaesser H.-P.;
 RT "A human intracellular apyrase-like protein, LALP70, localizes to
 RT lysosomal/autophagic vacuoles.";
 RL J. Cell Sci. 113:2473-2484(1999).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND ALTERNATIVE SPLICING.
 RX MEDLINE=20317099; PubMed=10856452;
 RA Biededick A., Kossan C., Kunz J., Elsaesser H.-P.;
 RT "First apyrase splice variants have different enzymatic properties.";
 RL J. Biol. Chem. 275:19018-19024(2000).
 RN [4]
 RP SEQUENCE OF 59-616 FROM N.A. (ISOFORM 2).
 RL DNA Res. 4:151-150(1997).
 CC -1- FUNCTION: Hydrolyzes preferentially nucleoside 5'-diphosphates,
 CC nucleoside 5'-triphosphates are hydrolyzed only to a minor extent.
 CC The order of activity with different substrates is UDP > GDP =
 CC CDP = TDP, AMP, ADP, ATP and UMP are not substrates. Preferred
 CC substrates for isoform 2 are CTP, UDP, CDP, GTP AND GDP, while
 CC isoform 1 utilizes UTP and TTP.
 CC -1- CATALYTIC ACTIVITY: A nucleoside diphosphate + H(2)O = a
 CC nucleotide + phosphate.
 CC -1- COFACTOR: Isoform 1 requires calcium and to a lesser extent,
 CC magnesium. Isoform 2 is equally dependent on calcium and
 CC magnesium.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC Localized in the Golgi and autophagic vacuoles/lysosomes.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=LALP70;
 CC IsoId=Q9Y227-1; Sequence=Displayed;
 CC Name=2; Synonyms=LALP70V;
 CC IsoId=Q9Y227-2; Sequence=VSP_003614;
 CC -1- TISSUE SPECIFICITY: Ubiquitous. Highest expression in testis and
 CC lowest in bladder.
 CC -1- SIMILARITY: Belongs to the GDA / CD39 NTPase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).
 CC or send an email to license@sib-sib.ch.
 CC -----
 DR EMBL; AF016032; AAC17217.1; -;
 DR EMBL; AJ013358; CAB40415.1; -;
 DR EMBL; AJ246168; CAB45685.1; JOINED.
 DR EMBL; AJ246166; CAB45685.1; JOINED.
 DR EMBL; AJ246167; CAB45685.1; JOINED.
 DR EMBL; AJ246168; CAB45685.1; JOINED.
 DR EMBL; AJ246169; CAB45685.1; JOINED.
 DR EMBL; AJ246170; CAB45685.1; JOINED.
 DR EMBL; AJ246171; CAB45685.1; JOINED.
 DR EMBL; AJ246172; CAB45685.1; JOINED.
 DR EMBL; AJ246173; CAB45685.1; JOINED.
 DR EMBL; AJ246174; CAB45685.1; JOINED.
 DR EMBL; AJ246175; CAB45685.1; JOINED.
 DR EMBL; AJ246176; CAB45685.1; JOINED.
 DR EMBL; AB002390; BAA21575.1; -;
 DR GeneW; HGNC:14573; LYSAL1.
 DR MIM; 607577; -;
 DR GO; GO:0030173; C:integral to Golgi membrane; IDA.

Query Match	Best Local Similarity	Score	D3 1,	Length	616;
Matches 132;	Conservative	69;	Mismatches 207;	Indels 123;	Gaps 20;
Dr GO; GO:0045134; P:uridine diphosphatase activity; IDA.					
Dr GO; GO:0006256; P:UDP catabolism; IDA.					
Dr InterPro; IPRO00407; GDAI CD33 _nnpase.					
Dr Pfam; PF01150; GDAI CD33; 1.					
Dr PROSITE; PS01238; GDAI_CD33_NNPASE; FALSE NEG.					
Kw HydroLase; Transmembrane; Glycoprotein; Calcium; Magnesium;					
Kw Alternative splicing; Golgi stack; Lysosome.					
FT DOMAIN 1 33					
FT TRANSMEM 34 54					
FT DOMAIN 55 559					
FT TRANSMEM 560 580					
FT DOMAIN 581 616					
FT CARBOHYD 404 404					
FT CARBOHYD 407 407					
FT VARSPIC 287 294					
FT CONFLICT 43 43					
FT SEQ SEQUENCE 616 AA; 70255 MW; 68A7EF7B3CB60544 CRC64;					
Query Match	13.6%;	Score 321.5;	D3 1,	Length 616;	
Best Local Similarity	24.9%;	Pred. No. 9.3e-18;			
Matches 132;	Conservative	69;	Mismatches 207;	Indels 123;	Gaps 20;
Dr GO; GO:0045134; P:uridine diphosphatase activity; IDA.					
Dr GO; GO:0006256; P:UDP catabolism; IDA.					
Dr InterPro; IPRO00407; GDAI CD33 _nnpase.					
Dr Pfam; PF01150; GDAI CD33; 1.					
Dr PROSITE; PS01238; GDAI_CD33_NNPASE; FALSE NEG.					
Kw HydroLase; Transmembrane; Glycoprotein; Calcium; Magnesium;					
Kw Alternative splicing; Golgi stack; Lysosome.					
FT DOMAIN 1 33					
FT TRANSMEM 34 54					
FT DOMAIN 55 559					
FT TRANSMEM 560 580					
FT DOMAIN 581 616					
FT CARBOHYD 404 404					
FT CARBOHYD 407 407					
FT VARSPIC 287 294					
FT CONFLICT 43 43					
FT SEQ SEQUENCE 616 AA; 70255 MW; 68A7EF7B3CB60544 CRC64;					
Query Match	13.6%;	Score 321.5;	D3 1,	Length 616;	
Best Local Similarity	24.9%;	Pred. No. 9.3e-18;			
Matches 132;	Conservative	69;	Mismatches 207;	Indels 123;	Gaps 20;
Dr GO; GO:0045134; P:uridine diphosphatase activity; IDA.					
Dr GO; GO:0006256; P:UDP catabolism; IDA.					
Dr InterPro; IPRO00407; GDAI CD33 _nnpase.					
Dr Pfam; PF01150; GDAI CD33; 1.					
Dr PROSITE; PS01238; GDAI_CD33_NNPASE; FALSE NEG.					
Kw HydroLase; Transmembrane; Glycoprotein; Calcium; Magnesium;					
Kw Alternative splicing; Golgi stack; Lysosome.					
FT DOMAIN 1 33					
FT TRANSMEM 34 54					
FT DOMAIN 55 559					
FT TRANSMEM 560 580					
FT DOMAIN 581 616					
FT CARBOHYD 404 404					
FT CARBOHYD 407 407					
FT VARSPIC 287 294					
FT CONFLICT 43 43					
FT SEQ SEQUENCE 616 AA; 70255 MW; 68A7EF7B3CB60544 CRC64;					
Query Match	13.6%;	Score 321.5;	D3 1,	Length 616;	
Best Local Similarity	24.9%;	Pred. No. 9.3e-18;			
Matches 132;	Conservative	69;	Mismatches 207;	Indels 123;	Gaps 20;
Dr GO; GO:0045134; P:uridine diphosphatase activity; IDA.					
Dr GO; GO:0006256; P:UDP catabolism; IDA.					
Dr InterPro; IPRO00407; GDAI CD33 _nnpase.					
Dr Pfam; PF01150; GDAI CD33; 1.					
Dr PROSITE; PS01238; GDAI_CD33_NNPASE; FALSE NEG.					
Kw HydroLase; Transmembrane; Glycoprotein; Calcium; Magnesium;					
Kw Alternative splicing; Golgi stack; Lysosome.					
FT DOMAIN 1 33					
FT TRANSMEM 34 54					
FT DOMAIN 55 559					
FT TRANSMEM 560 580					
FT DOMAIN 581 616					
FT CARBOHYD 404 404					
FT CARBOHYD 407 407					
FT VARSPIC 287 294					
FT CONFLICT 43 43					
FT SEQ SEQUENCE 616 AA; 70255 MW; 68A7EF7B3CB60544 CRC64;					
Query Match	13.6%;	Score 321.5;	D3 1,	Length 616;	
Best Local Similarity	24.9%;	Pred. No. 9.3e-18;			
Matches 132;	Conservative	69;	Mismatches 207;	Indels 123;	Gaps 20;
Dr GO; GO:0045134; P:uridine diphosphatase activity; IDA.					
Dr GO; GO:0006256; P:					

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=C57BL/6J; TISSUE=Lung;
RX MEDLINE=21085660; PubMed=11217851;
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Kuehl P., Lewis S., Matsuo Y., Gissi C., King B., Kochiwa H.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bulc C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberti P.,
RA Nordore P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Togo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Brain, and Mammary Gland;
RX MEDLINE=22398257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.C., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
RX MEDLINE=20317099; PubMed=10858452;
RA Biederbeck A., Kusan C., Kunz J., Elsaesser H.-P.;
RT "First apyrase splice variants have different enzymatic properties.";
RL J. Biol. Chem. 275:19018-19024(2000).
CC -!- FUNCTION: Hydrolyzes preferentially nucleoside 5'-diphosphates,
CC nucleoside 5'-triphosphates are hydrolyzed only to a minor extent
CC (by similarity).
CC -!- CATALYTIC ACTIVITY: A nucleoside diphosphate + H(2)O = a
CC nucleotide + phosphate.
CC -!- COFACTOR: Requires calcium and magnesium.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC Localizes in the Golgi and autophagic vacuoles/lysosomes (By
CC similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=LALP70;
CC IsoId=Q9DBT4-1; Sequences=Displayed;
CC Name=2; Synonyms=LALP70V;
CC IsoId=Q9DBT4-2; Sequences=VSP 003615;
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: Belongs to the GDAL / CD39 NTPase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AK004761; BAB23542.1; -;
DR EMBL; BC006924; AAH06924.1; -;
DR EMBL; BC043134; AAH43134.1; -;
DR MGD; MGI:1914714; Lysal1.
DR InterPro; IPR000407; GDAL CD39_NTPase.
DR Pfam; PF01150; GDAL CD39_1.
DR PROSITE; PS01238; GDAL CD39_NTPase; FALSE NEG.
KW Hydrolase; transmembrane; Glycoprotein; Calcium; Magnesium;
KW Alternative splicing; Golgi stack; Lysosome.
FT DOMAIN 1 33 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 34 54 POTENTIAL.
FT DOMAIN 55 559 LUMENAL (POTENTIAL).
FT TRANSMEM 560 580 POTENTIAL.
FT DOMAIN 581 613 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 407 407 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSELIC 287 294 Missing (in isoform 2).
FT /FTId=VSP 003615.
SQ SEQUENCE 613 AA; 69745 MW; DES28F512ABEF52F CRC64;
Query Match 13.6%; Score 320.5; DB 1; Length 613;
Best Local Similarity 25.8%; Pred. No. 1.1e-17;
Matches 120; Conservative 63; Mismatches 182; Indels 103; Gaps 17;
Qy 65 TAADGHEVYGFIMFDAGSTGTTRVHVQFTR--PPRETTLTHETFK-----AVKPGI 114
Db 80 TDITNFSVNYGIIVVDCGSSGRIFVYCWPHNGNPHDLDIRQMRDKNRKPVVMKIPGI 139
Qy 115 SAYADDERSAQOIRRELLDVAKQIDPFWKATPLVVKATAGLRLPGEAKKLOK-VK 173
Db 140 SEFATSPKVSVDYISPLLSFAASHVPRAKHETPLYLCTAGMRVLFPESQOKAILSDLLT 199
Qy 174 EVFKASPFVLGGDCVIMNGTDEGSAWITINPLTG-----SLKTPGGSS-- 218
Db 200 DIPHYVDFLPSDSHAENVISQKQGVAVICINFLGRFEHIEDDEAVVEVNIPEGSESE 259
Qy 219 -----VGMLDGGSGSTQIAP-LPVEGTQASPPGYL-TALRMFN-----RTYK 260
Db 260 AIVKRTAGVLDMGVSTQIAYEVPQTQVSPASSQOEVAQNLLAEFNLGCDVHQTEHYVR 319
Qy 261 LYSYSYIGLGLMSARL-----AIIIGGVEGQPAKDGKELVSPCLSPSPKGEW 306
Db 320 VYVATFLGFGNAARQRYEDRLFPASTVQNRLLGKQTG--LTPDAPLLPCLPLDKDEI 377
Qy 307 EHAETVYRVSGQKAAASLHLCARVSEVLQNRVHRTVEV-----KHVDYF 352
Db 378 QQNGQTLYLQG---TGDFDLC---RETLQPPMNKTNETQTSINGVYQPIHFQNSEFY 429
Qy 353 AFSYVYDLAAGVLIDAEKGSVLVGDPEIANKYVCT-----LETQPS 397
Db 430 GFSEFYCTEDV-----LRMGGDVNAARFTQAQKCATKWSILRRERDFRGLYASHADLHR 485
Qy 398 SPFSCMDLTVV-SLLLQEFQFPFRS-KVLKTRKIDNVETSWALGAIHF 443
Db 486 LKQCQKSAWMEVPHKGSFPVYTKNLTALQVYDKEVQWTLGAILY 533
RESULT 13
YBU4 CAEEL
ID YBU4 CAEEL STANDARD; PRT; 552 AA.
AC Q21815;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 16 16 C -> W (IN REF. 2).
FT CONFLICT 21 21 I -> G (IN REF. 1; AA SEQUENCE).
SQ SEQUENCE 493 AA; 54034 MW; F14FF4C3AA2F3603 CRC64;

Query Match
Best Local Similarity 13.3%; Score 315.5; DB 1; Length 493;
Matches 114; Conservative 67; Mismatches 180; Indels 69; Gaps 15;

QY 65 TRADGHEVF-----YGINFDAGSGTGVHVHVFQTRPPREPTLTH--ETKAVKPGLSA 116
DB 26 SAVDVKDFPPTGTYGLVDFDAGSTHLYVYQNPADKENGCTGVSVQVSECTVNGSGISS 85

QY 117 YADDVEKSAQGRRELLDVAQDIPDFWFKATPLVLKATAGLRLLPGEKAQKLLQKVKEVF 176
DB 86 YADDPAGAGASLKPCLDKAMAVIPVEQWQPTVYLGATAGWRLLEQNSTAEQVFAEVS 145

QY 177 KA-SPLVGDCCVSINMGTDGVSAMITNFI-----TGSUKTPGGSSV-GMLDLG 225
DB 146 KAIREFPVDFRGAQILTKGNEGSFGWITVNYLLETKIFSPAGKWEHPQNTVELGALDLG 205

QY 226 GGSTOIAFLPRVEGTLQASPPGYLTALSMENRTKLYSVSYLGLGLMSARLAILGGVEGQ 285
DB 206 GASTQITFGPV--TIEDKNTSVL--FRLYGTNSLYTHSVLCYQIQASRLMAALH-- 259

QY 286 PAKDG---KELVSPCLSPSFKGEWHAEV-----TYRVSGOKAAASL 324
DB 260 --QDSYVQVNIHPCVPKGYRIITABIYDPSVPTPSMLSPAQILTVTGTGNPAA-- 314

QY 325 HELCAARVSEVLQ-----NVRHTEEVKHV-----DFYAFSVYVYDLAAGVGLIDAEKGS 375
DB 315 ---CPTAILKLFNLTCGNARTCGDGVQPPVVRGQFFAPAGFYTFPLNLTGQOQSLSHV 371

QY 376 VVGDEFLAAKYCYRTLETQPOSS-----PFSQMDLTVVSLLLQERGFPRS--KVLKLRKI 429
DB 372 NATVWDFCNKWSSELVEFPQKEHLHYCVVGLVILLVLDVGYKDFBHTSNHFSQKA 431

QY 430 DNVEVSWALG 439
DB 432 GNADIGWTIG 441

RESULT 15
ENP3_HUMAN STANDARD; PRT; 529 AA.
AC O75355; O60495;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 3 (EC 3.6.1.5)
DE (NTPDase3) (Ecto-ATP diphosphohydrolase) (ATPDase) (Ecto-apyrase)
DE (CD39 antigen-like 3) (HB6).
GN ENTPD3 OR CD39L3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retinocytes;
RX MEDLINE=98341119; PubMed=9676430;
RA Chadwick B.P., Frischauf A.-M.;
RT "The CD39-like gene family: identification of three new human members
RT (CD39L2, CD39L3, and CD39L4), their murine homologues, and a member of
RT the gene family from Drosophila melanogaster.";
RL Genomics 50:357-367(1998).
RN [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=98342144; PubMed=9675246;

RA Smith T.M., Kirley T.L.;
RT "Cloning, sequencing, and expression of a human brain ecto-apyrase
RT related to both the ecto-ATPases and CD39 ecto-apyrases.";
RL Biochim. Biophys. Acta 1386:65-78(1998).
RN [3]
RP REVISIONS.
RA Smith T.M., Kirley T.L.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP MUTAGENESIS OF TRP-187; ASP-219 AND TRP-459.
RX MEDLINE=99249795; PubMed=10231536;
RA Smith T.M., Lewis Carl S.A., Kirley T.L.;
RT "Mutagenesis of two conserved tryptophan residues of the E-type
RT ATPases: inactivation and conversion of an ecto-apyrase to an
RT ecto-NTPase.";
RL Biochemistry 38:5849-5857(1999).
RN [5]
RP MUTAGENESIS OF ARG-67; ARG-143; ARG-146; GLU-182; ASN-191; SER-224 AND
RP GLN-226.
RX MEDLINE=21197753; PubMed=11300774;
RA Yang F., Hicks-Berger C.A., Smith T.M., Kirley T.L.;
RT "Site-directed mutagenesis of human nucleoside triphosphate
RT diphosphohydrolase 3: the importance of residues in the apyrase
RT conserved regions.";
RL Biochemistry 40:3943-3950(2001).
CC -I- FUNCTION: Has a threefold preference for the hydrolysis of ATP
CC over ADP.
CC -I- CATALYTIC ACTIVITY: ATP + 2 H(2)O = AMP + 2 phosphate.
CC -I- COFACTOR: Requires calcium and magnesium.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -I- TISSUE SPECIFICITY: Expressed in adult brain, pancreas, spleen and
CC prostate. Moderate or low expression is seen in most tissues. Not
CC expressed in liver and peripheral blood leukocytes.
CC -I- SIMILARITY: Belongs to the GDAL / CD39 NTPase family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC ENBL; AF039917; AAC39884.1; -;
CC EMBL; AF034840; AAC09236.2; -;
CC Genew; HGNC:3365; ENTPD3.
CC MIM; 603161; -;
CC InterPro; IPR000407; GDAL_CD39_NTPase.
CC Pfam; PF01150; GDAL_CD39; 1.
CC PROSITE; PS01238; GDAL_CD39_NTPase; 1.
KW Hydrolyase; Transmembrane; Glycoprotein; Calcium; Magnesium.
FT DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 23 43 POTENTIAL.
FT DOMAIN 44 48 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 486 506 POTENTIAL.
FT DOMAIN 507 529 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 454 454 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT MUTAGEN 67 67 R->G: INCREASE OF ACTIVITY.
FT MUTAGEN 143 143 R->A: LOSS OF ACTIVITY.
FT MUTAGEN 143 143 R->K: INCREASE OF ACTIVITY.
FT MUTAGEN 146 146 R->N: NO EFFECT.
FT MUTAGEN 146 146 R->P: INCREASE OF ATPASE ACTIVITY,
FT MUTAGEN 146 146 DECREASE OF ADPASE ACTIVITY.
FT MUTAGEN 146 146 R->T: INCREASE OF ACTIVITY.
FT MUTAGEN 182 182 R->D: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 182 182 E->Q: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 187 187 W->A: COMPLETE LOSS OF ACTIVITY.

```
FT MUTAGEN 191 191 N->A: LOSS OF ATPASE ACTIVITY, INCREASE
FT MUTAGEN 219 219 OF ADPASE ACTIVITY.
FT MUTAGEN 224 224 D->E: INCREASE OF ACTIVITY.
FT MUTAGEN 226 226 S->A: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 459 459 Q->A: LOSS OF ACTIVITY.
FT MUTAGEN 459 459 W->A: INCREASE OF ACTIVITY, ESPECIALLY
THE ATP HYDROLYSIS.
FT CONFLICT 496 496 V->A (IN REF. 3).
SQ SEQUENCE 529 AA; 59133 MM; 505D2PEC12978B88 CRC64;
```

Query Match 13.3%; Score 314.5; DB 1; Length 529;
Best Local Similarity 25.3%; Pred. No. 2.7e-17;
Matches 113; Conservative 76; Mismatches 166; Indels 91; Gaps 20;

```
QY 74 YGIMPDASTGTGRHVVFQTRPPREPTLTHETK-AVK-PGLSAYADVEKSAQGIREL 131
Db 57 YGIVLDASSRTYVYQWPAKENTNGVSOQTKSVKSGISSYGNPDVPRAFEEC 116
QY 132 LDVAKQDIPDFWFKATPLVLKATAGRLLPGEK--AQKLIQKVEVKASPPFLVGDGDC- 187
Db 117 MOKYKGVPSHLHSGTPIHLGATAGMRLLRLQNETANEVLESIQSYFKSQPF---DFR 172
QY 188 -VSIMNGTDEGVSAWITINFLTGS-----LKTPEGSSVGMLDIGGSGTQIAFLPR 236
Db 173 GAOIISGQEGEGYGTANYLMGNFLEKRLMHWVHPHGVETTGALDLGASQISFVAG 232
QY 237 VEGTLQASPPGYLTALRMENRYKLXSYSLGLGMSAR---LAILGVEGQPAKDGKEL 293
Db 233 EKMDLNTSD---IMOVSLXGYVYTLTYHSFOCYGRNEAKKFLMNL--LQNSPTKN--HL 285
QY 294 VSPCLSPSFKGEWEHAEV-----TYRVSGQKAAASLHELCAARVS 333
Db 286 TNPCTPRDYISIFTWGHVDSLCTVDQRPESYNPDVITFEGTDPS-----LCKEKVA 339
QY 334 EVLQWRVHRTREVKHVD-----FYASYYDLAAGVGLDAEKGSLVYGDPEIA 383
Db 340 SIIDFKACHDQETCSFDGVYQPKIKGPVAFAGFYTASALNL-----SGSFSLDITNSS 394
QY 384 AKYVCRTLETO-POSSP-----FSCMDLTYV-SLLDEFGPPRS--KYLKLTREKIDN 431
Db 395 TWNFGSQWWSQPLPLPKFEDEVYARSYCFSANITYHLFVNGYKFTETWPOIHFEKEVGN 454
QY 432 VETSWALGAIPIHYIDSLNRQ--KSP 454
Db 455 SSIAWSLIG---YMLSLTNQIPASBP 476
```

Search completed: June 8, 2004, 09:48:49
Job time : 20 secs


```
Db 148 VEKSAQGIKRELLDVAQKQDIPDFWKATPLVLKATAGLRLLPGEKAQKLLQKVKEVFKASP 207
QY 181 FLVGDDCVSIMNGTDEGVSAWITINFLTGLSKTPGGSSVGMGLDLGGSTQIAFLPRVEGT 240
Db 208 FLVGDDCVSIMNGTDEGVSAWITINFLTGLSKTPGGSSVGMGLDLGGSTQIAFLPRVEGT 267
QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGGVEGPAKDGKELVSPCLSP 300
Db 268 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGGVEGPAKDGKELVSPCLSP 327
QY 301 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVKHVDVFAFSYYIDL 360
Db 328 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVKHVDVFAFSYYIDL 387
QY 361 AAGVGLIDAEKGSLLVVGDFEIAAKVVCRTLETQPOSSPSCMDLTYVSLLLQEFGRPS 420
Db 388 AAGVGLIDAEKGSLLVVGDFEIAAKVVCRTLETQPOSSPSCMDLTYVSLLLQEFGRPS 447
QY 421 KVLKLRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456
Db 448 KVLKLRKIDNVETSWALGAIFHYIDSLNRQKSPAS 483

RESULT 2
Q725B5 PRELIMINARY; PRT; 484 AA.
AC Q725B5;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DE CD39L2 nucleotide.
GN ENTPD6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Ivanenkov V.V., Murphy-Piedmonte D.M., Kirley T.L.;
RT "Bacterial Expression, Characterization, and Disulfide Bond
RT Determination of Soluble Human NTPase6 (CD39L2) Nucleotidase:
RT Implications for Structure and Function.";
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR ENBL; AY327581; AAP92131.1; -.
SQ SEQUENCE 484 AA; 53274 MW; D00A5D2915DF36CE CRC64;

Query Match 99.9%; Score 2361; DB 4; Length 484;
Best Local Similarity 99.8%; Pred. No. 6.9e-196;
Matches 455; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKISHGSLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFSITRAAPGARWGOQAH 60
Db 29 MRKISHGSLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFSITRAAPGARWGOQAH 88
QY 61 SPLGTAADGHEVYIGIMFDAGSTGTGVHVFOQTRPREPTLTHTFKAQKGLSAYADD 120
Db 89 SPLGTAADGHEVYIGIMFDAGSTGTGVHVFOQTRPREPTLTHTFKAQKGLSAYADD 148
QY 121 VEKSAQGIKRELLDVAQKQDIPDFWKATPLVLKATAGLRLLPGEKAQKLLQKVKEVFKASP 180
Db 149 VEKSAQGIKRELLDVAQKQDIPDFWKATPLVLKATAGLRLLPGEKAQKLLQKVKEVFKASP 208
QY 181 FLVGDDCVSIMNGTDEGVSAWITINFLTGLSKTPGGSSVGMGLDLGGSTQIAFLPRVEGT 240
Db 209 FLVGDDCVSIMNGTDEGVSAWITINFLTGLSKTPGGSSVGMGLDLGGSTQIAFLPRVEGT 268
QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGGVEGPAKDGKELVSPCLSP 300
Db 269 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGGVEGPAKDGKELVSPCLSP 328
QY 301 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVKHVDVFAFSYYIDL 360
```

```
Db 329 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVKHVDVFAFSYYIDL 388
QY 361 AAGVGLIDAEKGSLLVVGDFEIAAKVVCRTLETQPOSSPSCMDLTYVSLLLQEFGRPS 420
Db 389 AAGVGLIDAEKGSLLVVGDFEIAAKVVCRTLETQPOSSPSCMDLTYVSLLLQEFGRPS 448
QY 421 KVLKLRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456
Db 449 KVLKLRKIDNVETSWALGAIFHYIDSLNRQKSPAS 484

RESULT 3
Q8N3H3 PRELIMINARY; PRT; 503 AA.
AC Q8N3H3;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein (fragment).
GN DKFZP761J1915.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Amalgam;
RA Ansgorge W., Wirkner U., Mewes H.W., Weill B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR ENBL; AL834158; CAD38864.1; -.
DR InterPro; IPR000407; GDAI_CD39_NTPase.
DR Pfam; PF01150; GDAI_CD39; 1.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 503 AA; 54763 MW; 163933P9139D89F CRC64;

Query Match 99.8%; Score 2361; DB 4; Length 503;
Best Local Similarity 99.8%; Pred. No. 7.3e-196;
Matches 455; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKISHGSLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFSITRAAPGARWGOQAH 60
Db 48 MRKISHGSLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFSITRAAPGARWGOQAH 107
QY 61 SPLGTAADGHEVYIGIMFDAGSTGTGVHVFOQTRPREPTLTHTFKAQKGLSAYADD 120
Db 108 SPLGTAADGHEVYIGIMFDAGSTGTGVHVFOQTRPREPTLTHTFKAQKGLSAYADD 167
QY 121 VEKSAQGIKRELLDVAQKQDIPDFWKATPLVLKATAGLRLLPGEKAQKLLQKVKEVFKASP 180
Db 168 VEKSAQGIKRELLDVAQKQDIPDFWKATPLVLKATAGLRLLPGEKAQKLLQKVKEVFKASP 227
QY 181 FLVGDDCVSIMNGTDEGVSAWITINFLTGLSKTPGGSSVGMGLDLGGSTQIAFLPRVEGT 240
Db 228 FLVGDDCVSIMNGTDEGVSAWITINFLTGLSKTPGGSSVGMGLDLGGSTQIAFLPRVEGT 287
QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGGVEGPAKDGKELVSPCLSP 300
Db 288 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGGVEGPAKDGKELVSPCLSP 347
QY 301 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVKHVDVFAFSYYIDL 360
Db 348 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVKHVDVFAFSYYIDL 407
QY 361 AAGVGLIDAEKGSLLVVGDFEIAAKVVCRTLETQPOSSPSCMDLTYVSLLLQEFGRPS 420
Db 408 AAGVGLIDAEKGSLLVVGDFEIAAKVVCRTLETQPOSSPSCMDLTYVSLLLQEFGRPS 467
QY 421 KVLKLRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456
Db 468 KVLKLRKIDNVETSWALGAIFHYIDSLNRQKSPAS 503
```

```

RESULT 4
Q8CH23 PRELIMINARY; PRT; 278 AA.
AC Q8CH23;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Similar to ectonucleoside triphosphate diphosphohydrolase 6.
GN ENTPD6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RA Strauberg R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR MGD; B038126; A038126.1; -.
DR MGD; MGI:1202295; Ectpd6.
DR GO; GO:0016787; F:hydrolase activity, IEA.
DR InterPro; IPR00407; GDAI_CD39_NTPase.
DR Pfam; PF01150; GDAI_CD39; 1.
KM Hydrolase.
SQ SEQUENCE 278 AA; 30130 MW; 75A92DD1AC76297F CRC64;

Query Match 52.0%; Score 1229.5; DB 11; Length 278;
Best Local Similarity 85.6%; Pred. No. 4e-98;
Matches 237; Conservative 17; Mismatches 18; Indels 5; Gaps 1;

QY 1 MRKISNGSLRVAFAVAVPLGCVGFYVAVYIKHRAATATQAFSSITRAPAGARWGOAH 60
DB 1 MRKIPNNGTLMRTKATVAPLGLCVGLFTYVAYIKHRAAQAFTTLAGASGARWGOAH 60

QY 61 SPLGTADGHEVFGYGVNPDAGSTGTRVHVPQFTRPREPTLTTHETFKAVKPGLSAYAD 120
DB 61 SSPGSAARGHEVFYGVNPDAGSTGTRVHVPQFTRPREPTLTTHETFKAVKPGLSAYAD 120

QY 121 VEKSAOGIRELDVAKODIPDFPKATPLVAKATAGRLGPKAKOKLQKVEVFKASP 180
DB 121 VEKSAOGIRELDVAKODIPDFPKATPLVAKATAGRLGPKAKOKLQKVEVFKASP 180

QY 121 VEKSAOGIRELDVAKODIPDFPKATPLVAKATAGRLGPKAKOKLQKVEVFKASP 180
DB 121 VEKSAOGIRELDVAKODIPDFPKATPLVAKATAGRLGPKAKOKLQKVEVFKASP 180

QY 181 FLVGDGVSINMGDEGVSAMITINFLTGLSKTPGSSVGMLDLGGGSTOIAELPRVEGT 240
DB 181 FLVGDGVSINMGDEGVSAMITINFLTGLSKTPGSSVGMLDLGGGSTOIAELPRVEGT 240

QY 241 LQASPPGYLTLMKMFRTTYKLYSYTYGLGMSARLA 277
DB 241 LQASPPGYLTLMKMFRTTYKLYSYTYGLGMSARLA 277

RESULT 5
Q8CD29 PRELIMINARY; PRT; 427 AA.
AC Q8CD29;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 5.
GN ENTPD5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).

```

```

DR EMBL; AK031581; BAC27461.1; -.
DR MGD; MGI:1321385; Entpd5.
DR InterPro; IPR00407; GDAI_CD39_NTPase.
DR Pfam; PF01150; GDAI_CD39; 1.
SQ SEQUENCE 427 AA; 47101 MW; 6E3773C842B58477 CRC64;

Query Match 42.6%; Score 1007; DB 11; Length 427;
Best Local Similarity 50.0%; Pred. No. 1.6e-78;
Matches 213; Conservative 62; Mismatches 133; Indels 18; Gaps 6;

QY 37 ATATQAFPSITRAAPGA-----RMGO-----QAHSEPLGTADGHEVFGYGVNPDAGSTG 84
DB 2 ATSWGAVVMLLIACGSGVFYREQQTWEQVFLSMCINVSAG---TFYGVNPDAGSTG 58

QY 85 TRVAVFQRT-RPREPTPLTTHETKAVKPGLSAYADVEKSAOGIRELDVAKODIPDF 143
DB 59 TRHYVTFVQKTAGQGLPFLBGEIIPDSVKGSAFVDPKQAGETVQELLEVAKOSIPRSH 118

QY 144 WKATPLVAKATAGRLGPKAKOKLQKVEVFKASPFLVGDGVSINMGDEGVSAMIT 203
DB 119 WERTVVKATRGRLPLPEQKQALLLEVEBIFKNSPFLVDGVSIMDSYEGILAMVT 178

QY 204 INFLTGLSKTPGSSVGMLDLGGGSTOIAELPRVEGTQASPPGYLTLMKMFRTTYKLYS 263
DB 179 VNFITGLHGRGQETVGTLDLGASTQITFLPQFEKTEKTPRGYLTSEFMFNSTFKLYT 238

QY 264 VSYLGLGMSARIALTGVGEGPAKDKELVSPCLSPFKEWEMHAEVTVSGOKAAS 323
DB 239 HSYLDFGKAAPALATLGLALENK-GTDGTFPSACLPKMLEMWFYGVKYIYGQNGSGEM 297

QY 324 LHELCAAVSEVLTQNRVHRTVEVKHVDFAFSYYDDLAAGVGLDAEKGSLVVDPEFA 383
DB 296 GEPGCVAVLRVQKQKHQPEVSGSAFYASYDDRADTHLIDYEGVLYKVEDPERK 357

QY 384 AKYVKTLETQDQSSPESCMILTYSLLOE-FGPPRKVKLTIRKIDNVETSMALGIF 442
DB 358 AREVCDNIGSPSSGSPFLCMDLTYITALLKQGFADGTLQTLTKVNIETGALGATF 417

QY 443 HYDSL 448
DB 418 HLIQSL 423

RESULT 6
Q8BR23 PRELIMINARY; PRT; 427 AA.
AC Q8BR23;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 5.
GN ENTPD5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RA MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR MGD; MGI:1321385; BAC32507.1; -.
DR InterPro; IPR00407; GDAI_CD39_NTPase.
DR Pfam; PF01150; GDAI_CD39; 1.
SQ SEQUENCE 427 AA; 47123 MW; 87BF2CC1CC1FCB9 CRC64;

Query Match 42.3%; Score 1000; DB 11; Length 427;
Best Local Similarity 49.8%; Pred. No. 6.5e-78;

```

Matches 212; Conservative 61; Mismatches 135; Indels 18; Gaps 6;

QY 37 ATATQAFSSITRAAPGA-----RWGO-----QAHSPLGTAAAGCHEVFYGMFDAGSTG 84
DB 2 ATSWGAVFLLIACVSTVFYREOQTWFGVFLSSMCPINVSAG---TFYGMFDAGSTG 58
QY 85 TRVHVFOFT-RPPRETPTLTHETFKAVKPLGSLAYADDVEKSAQGIREDLDVAKQIDPDF 143
DB 59 TRIHVTFVQKTAGQQLFFLEGEIFDSVKPGLSAFVDQPKQAETVQELLEVAKDSIPRSH 118
QY 144 WKATPLVLKATAGRLLPGEKAQKLOKVEFKASPFPLVGGDDCVSINMGTDGVSAMIT 203
DB 119 WERTPXKATAGRLLLPQKAQALLLEVEEIKFSPFLVPGSVSIMGSGVEGLAWVT 178
QY 204 INFLTSLKTPGSSVGMGLDGGSTQIAFLPRVGTLOASPYPGTLALRPNRTYKLS 263
DB 179 VNFLTQLHGRGOETVGLDGGASTQITFLPQFEKTLTQTPRGVLTSPFENSTFKLYT 238
QY 264 YSVLGLGLMSARLAIIGGVGEGQPAKDGKELVSPCLSPFKGEWEHAETVYRVSQKAAAS 323
DB 239 HSVLGFGLKAARLATLGALEAK-GTDTHTFRSACLPRWLEAEWIFGGVKYQYGGNQEGEM 297
QY 324 LHELCAARVSEVLQNRVHRTTEEVKHYVDFYAFSYYYDLAAGVLIDAEKGSVLVGGDFEIA 383
DB 298 GFEPCEAEVLRVVQGLKHQPEEVGRGSAFYAFSVYYDRAADTHLIDYKGGVLLKVEDFERK 357
QY 384 AKVVCETLTQPOSSPSCMDLTYVSLLOE-FGPRSKVLKTRKIDNVETSWALGAIF 442
DB 358 AREVCNLSGSSGSPFLCMDLTYITALLKDGFGFADGTLTQLTKVNNITGVALGATF 417
QY 443 HYIDSL 448
DB 418 HLLQSL 423

RESULT 7
Q96RXO PRELIMINARY; PRT; 428 AA.
AC Q96RXO; STRAIN=C57BL/6J; TISSUE=Thymus;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Pcpb proto-oncogene protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20173601; PubMed=10708485;
RA Recio J.A., Zambrano N., Pena, Ld, Reig J.A., Rhoads A., Rouzaut A.,
RA Notario V.;
RT "The human PCPH proto-oncogene: cDNA identification, primary
RT structure, chromosomal mapping, and expression in normal and tumor
RT cells";
RL Mol. Carcinog. 27:229-236(2000).
DR EMBL; AF136572; AAK82950.1; -;
DR InterPro; IPR000407; GDAI_CD39_NTPase.
DR Pfam; PF01150; GDAI_CD39_1.
SQ SEQUENCE 428 AA; 47431 MW; F2C4F7DE650A44F6 CRC64;

Query Match 42.0%; Score 992; DB 4; Length 428;
Best Local Similarity 52.2%; Pred. No. 3.2e-77;
Matches 203; Conservative 56; Mismatches 124; Indels 6; Gaps 4;

QY 62 PLGTAADGHEVFYGMFDAGSTGTRVHVFOFT-RPPRETPTLTHETFKAVKPLGSLAYADD 120
DB 40 PINVSA---STLYGINFDAGSTGTRHVHTVYVQKAAGQLPILGEVFSVPGLSAFVDQ 96
QY 121 VKSAQGIREDLDVAKQIDPDFFWKATPLVLKATAGRLLLPGEKAQKLOKVEFKASP 180
DB 97 PKQGAETVQGLLEVAKDSIPRSHWKKTPVVLKATAGRLLLPGEKAQKLOKVEFKASP 156

Matches 212; Conservative 61; Mismatches 135; Indels 18; Gaps 6;

QY 181 FLVDDCVSIMNGTDBGVSAWITINFLTSLKTPGSSVGMGLDGGSTQIAFLPRVEGT 240
DB 157 FLVPKGSVIMDGSDEGILLAWTVNFLTQLHGRHOETVGLDGGASTQITFLPQFEKT 216
QY 241 LQASPPGYLTALRPNRTYKLSYVSLGLGIMSAARLAIIGGVGEGQPAKDGKELVSPCLSP 300
DB 217 LEQTPRGVLTSPFENSTFKLYTHSVLGFGLKAARLATLGALETE-GTDTHTFRSACLPR 275
QY 301 SFGGEWEHAETVYRVSQKAAASLHELCAARVSEVLQNRVHRTTEEVKHYVDFYAFSYYYDL 360
DB 276 WLEAEWIFGGVKYQYGGNQEGVEGFPCEAEVLRVVGRKLHQPPEEVQSGFYAFSYYYDR 335
QY 361 AAGVGLIDAEKGSVLVGGDFEIAAKVVCETLTQPOSSPSCMDLTYVSLLOE-FGPR 419
DB 336 AVDTMDIDYKGGVLLKVEDFERKAREVCNLENFTSGSPFLCMDLTYITALLKDGFGFAD 395
QY 420 SKVLKTRKIDNVETSWALGAIFHYIDSL 448
DB 396 STVLQTLTKVNNITGVALGATFHLQSL 424

RESULT 8
Q8CDB6 PRELIMINARY; PRT; 224 AA.
AC Q8CDB6; STRAIN=C57BL/6J; TISSUE=Thymus;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 6.
GN ENTPD6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA THE PANTOM Consortium.
RA THE RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
DR EMBL; AK030789; BAC27138.1; -;
DR MGD; MGI:1202295; Entpd6.
DR InterPro; IPR000407; GDAI_CD39_NTPase.
DR Pfam; PF01150; GDAI_CD39_1.
SQ SEQUENCE 224 AA; 24538 MW; 305DCC373B5B4A8E CRC64;

Query Match 40.5%; Score 957; DB 11; Length 224;
Best Local Similarity 87.1%; Pred. No. 1.2e-74;
Matches 182; Conservative 12; Mismatches 15; Indels 0; Gaps 0;

QY 1 MRKISNHGSLRVAKVAPLGLCVGVETIYVAYIKWHEATATQAFPSITRAAPGARWQQA 60
DB 1 MRKIPNHGTLRMTKVAYPLGLCVGLIYVAYIKWHEASAAQAFITAGASGARWQQA 60
QY 61 SPLGTAADGHEVFYGMFDAGSTGTRVHVFOFT-RPPRETPTLTHETFKAVKPLGSLAYADD 120
DB 61 SPSGSAARGHEVFYGMFDAGSTSTRIHVFOFARPPGETPTLTHETFKALPKLSAYADD 120
QY 121 VKSAQGIREDLDVAKQIDPDFFWKATPLVLKATAGRLLLPGEKAQKLOKVEFKASP 180
DB 121 VKSAQGIREDLDVAKQIDPDFFWKATPLVLKATAGRLLLPGEKAQKLOKVEFKASP 180
QY 181 FLVDDCVSIMNGTDBGVSAWITINFLTG 209
DB 181 FLVDDCVSIMNGTDBGVSAWITINFLTG 209

RESULT 9
Q8WUB3 PRELIMINARY; PRT; 407 AA.
ID Q8WUB3

AC Q9WUB3; 01-MAR-2002 (TRENBLREL. 20, Created)
 DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)
 DT 01-MAR-2002 (TRENBLREL. 24, Last annotation update)
 DE Similar to econucleoside triphosphate diphosphohydrolase 5.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Tissue=Colon;
 RA Struysberg R;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC020966; AAH20966.1; -
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR000407; GDA1_CD39_NTPase.
 DR Pfam; PF01150; GDA1_CD39_1.
 KW Hydrolase.
 SQ SEQUENCE 407 AA; 45336 MW; D92A5F7DC9EC9E5B CRC64;
 Query Match 38.5%; Score 909; DB 4; Length 407;
 Best Local Similarity 51.8%; Pred. No. 4.7e-70;
 Matches 189; Conservative 51; Mismatches 119; Indels 6; Gaps 4;
 QY 62 PLGTADGHEVYVYMGFAGSTGTVHVFQFT-RRPREPTLTTHETFAVAKGSAVADD 120
 DB 40 PLVNA---STYGLMPDAGSTGTVHVFQFTGQGLPFLGEGVDFSDVSPGLSATVDQ 96
 QY 121 VEKSAQGIREDLDAVAKODIPDFFWKATPLVKATNGRLPGEKAKLQYKVEFKASP 180
 DB 97 PRQGEATYQGLLEVAKDISPRSHMKKTPLVKATNGRLPGEKAKLLEFVEKIFRSP 156
 QY 161 FLVGDGCSIMNGTDEGVSAMTINFLGSLKTPGSSVGMLDGGSSTQAFIPRVEGT 240
 DB 157 FLVPGSVSIMDGSDEGLAWTVFLTGQHGHRQETVGLDGGASTQITFLPQPEKT 216
 QY 241 IQASPPGYLTALRMNRTYKLYSYSLGLGMSARLALIGVEGQPAKGEIVSPCLSP 300
 DB 217 LEQTRGVLTSPENMNSTYKLYTHSYLFGKAKARLALGLAFTE-GRDGHTRFSACLR 275
 QY 301 SFKGEWEHAEVTVYVSGQKAAASLHELCAARYSEVLQNRKVRTEVYKAVDYAASYIDL 360
 DB 276 WLEAWMITGQVYQYQGNQGEVGFPEYAEVLRVWRKMLQPEBVGSGSYFASYYDR 335
 QY 361 AAGVGLIDAEKGSIVVGFELIAKYVCRLETQOSPFSCMDLTYSLLQF-FGPPR 419
 DB 336 AVDTMIDYERKGIKVEDFERKAREVCDNLENFTSGSPFLCMLSTYTLALKGFGFAD 395
 QY 420 SKVLK 424
 DB 396 STVLQ 400
 RESULT 10
 ID 076268 PRELIMINARY; PRT; 461 AA.
 AC 076268;
 DT 01-NOV-1998 (TRENBLREL. 08, Created)
 DT 01-NOV-1998 (TRENBLREL. 08, Last sequence update)
 DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
 DE NTPase protein (JDI1641P).
 GN NTPASE OR CG3059.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=2019606; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., McKillop G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Baller R.M., Basu A., Bakendale U., Bayraktiroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhattacharya D., Borker P.,
 RA Borkova D., Borkman M.R., Bouck J., Brokstein P., Brotshakov S.,
 RA Burris K.C., Busam D.A., Butler R., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Galburt M.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lesko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mlshina N.V., Moberg C., Morris D., Mostrel A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman K.S., Sanders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Teclor C., Turner R., Venter R., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers R.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98341119; PubMed=9676430,
 RA Chadwick B.P., Fritsch A.M.,
 RT "The CD39-like gene family: Identification of three new human members
 RT (CD3912, CD3913, and CD3914), their murine homologues, and a member of
 RT the gene family from Drosophila melanogaster.";
 RL Genomics 50:357-367(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise B., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda C.J.,
 RA Nuno J., Pacle J., Parag V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB003581; AAF51182.1; -
 DR EMBL; AF041048; AAC39133.1; -
 DR EMBL; AY061134; AAL28682.1; -
 DR FLYBase; FBgn0024947; NTPase.
 DR InterPro; IPR000407; GDA1_CD39_NTPase.
 DR Pfam; PF01150; GDA1_CD39_1.
 SQ SEQUENCE 461 AA; 50845 MW; 27D0031P91A9DD1 CRC64;
 Query Match 26.3%; Score 621; DB 5; Length 461;
 Best Local Similarity 39.9%; Pred. No. 5.6e-45;
 Matches 158; Conservative 60; Mismatches 142; Indels 35; Gaps 13;
 QY 71 EVFYGLMPDAGSTGTVHVFQFTRRPREPT-PLTHETFAVAKGSAVADDVEKSAQGR 129
 DB 76 KQYAAIIDAGSTGTVHVFQFTRRPREPT-PLTHETFAVAKGSAVADDVEKSAQGR 135
 QY 130 ELIDVAKODIPDFFWKATPLVKATNGRLPGEKAKLQYKVEFKASPPLVGDGCS 189
 DB 136 LLIDDEARAPITPREHWSSTPLVKATNGRLPGEKAKLQYKVEFKASPPLVGDGCS 195

DR InterPro: IPR002034; AIPM/HcIt synth.
 DR InterPro: IPR000407; GDAI_CD39_NTPase.
 DR Pfam: PF01150; GDAI_CD39_1.
 DR PROSITE: PS00815; AIPM_HOMOCIT_SYNTH_1, 1.
 DR SEQUENCE 479 AA; 53844 MW; 7EDC02A9D54A8ED CRC64;

Query Match 25.0%; Score 590.5; DB 5; Length 479;
 Best Local Similarity 34.2%; Pred. No. 2,6e-42;
 Matches 134; Conservative 78; Mismatches 155; Indels 25; Gaps 9;

QY 73 FYGIMDAGSTGRVAVFOGTRPRRPT-----TLHETPKAVKPGLSAYADVEKSAQG 127
 DB 42 FFIIVIDAGSTGRHLHYKFIHDPALASHGMPKVEKIFQEVKPGLSFAKSPSSAAS 101
 QY 128 IRELLDVAKODIPDFWKATPLVKATAGRLPGEKAQKLQKVE-VFKASPELVGD 186
 DB 102 LEPLLRARKEVHFHFMWEKTPILTKATAGRLPDMADILLESVERIFNSGFFAAPD 161
 QY 187 CVAINGTDEGVASMTINFLTGSJ-----KTPGSSVGMIDLGGSTQIAFLPRV 237
 DB 162 AVNMPGSDGVYSWFTLNTLLETLFTDEPTVGKPAHRSVAAPDLGGSTQITLTPMN 221
 QY 238 EGTLOASPGLYALMFMNRTYKLSYSYLGILMSARLAILGVGQPAKDGKELVSPC 297
 DB 222 EAVP-SHVGVEYEDIDFFGHIRLFTHSFLGNGLINAKNLILQLETDNEIESTHQLITSC 280
 QY 298 LSPSEFK-GEWEHAEVTVRSVQKAAASLHELCAARVSEVLQ-NRVHTEBVKXVDPAFS 355
 DB 281 MPEGYQLTEWEYALKFNNING--SSHSFESCYGTTNPFVSESEIHMLRELKSGPYLFS 338
 QY 356 YYVDLAGVGLDAEKGSLVGDPELIAKYVCRTELTQGS-----PSCMDLTVV-SL 410
 DB 339 YFPRLANSGLVKNEGSKTELRFKAETACRRETEIDDSGHMMPWOCIDLTYIVSL 398
 QY 411 LLOEFGPRSKVLKTRKIDNVETSMALGAIF 442
 DB 399 LRQGYQFEDNQPLVLAKKIKMEVSWGGLAF 430

RESULT 13
 Q9HEM6 PRELIMINARY; PRT; 522 AA.
 AC Q9HEM6;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Guanosine diphosphatase.
 GN GDAI.
 OS Kluveromyces lactis (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluveromyces.
 NCBI_TaxID=28985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MGI/2;
 RX MEDLINE=21319012; PubMed=11425802;
 RA Lopez-Avalos M.D., Uccelletti D., Abeijon C., Hirschberg C.B.;
 RT "The UDPase activity of the Kluveromyces lactis Golgi GDPase has a
 role in uridine nucleotide sugar transport into Golgi vesicles.";
 RL Glycobiology 11:413-422(2001).
 DR EMBL; AJ401304; CAC21576.1; -;
 DR InterPro: IPR000407; GDAI_CD39_NTPase.
 DR Pfam; PF01150; GDAI_CD39_1.
 DR PROSITE; PS01238; GDAI_CD39_NTPase; 1.
 DR PS01238; GDAI_CD39_NTPase; 1.
 SQ SEQUENCE 522 AA; 56863 MW; A7A27902607A6732 CRC64;

Query Match 22.5%; Score 531.5; DB 3; Length 522;
 Best Local Similarity 34.4%; Pred. No. 4e-37;
 Matches 153; Conservative 63; Mismatches 142; Indels 87; Gaps 19;
 QY 68 DGHEVYFYGIMPDAAGTGRVAVFOF--TRPREPTLTHETPKAVKPGLSAYADVEKSA 124
 DB 93 EGHK--YVVMIDAGSTGRVAVVEFDVCTG---PILINTEFMKPKGLSSPFDIVAGA 146

QY 125 AAGIRELLDVAKODIPDFWKATPLVKATAGRLPGEKAQKLQKVE-KVFKASPELV 183
 DB 147 AKSIDPLKILAMDVAVPDKXKNCFTPVANKATAGRLMDDEKSSKILAQVRHLEGDYFPV 206
 QY 184 GD-DCVINGTDEGVASMTINFLTGSJ-----KTPGSSVGMIDLGGSTQIAFLPRV 238
 DB 207 VDGSDVSMDEEGVAVAWYANILGNIGAGSLP--TAAYVDLGGSTQITLVEEP-- 260
 QY 229 GTLOASPGLYALMFMNRTYKLSYSYLGILMSARLAILGVGQPAKDGKELVSPC 297
 DB 261 ---SFPNEMKMDGEHKEKLSFGCHYTLQFSLHGLMGKNKINTELAVNAISSGT 316
 QY 282 -VEGQPAKDGKELVSPCISPFKGEWEHAEV-----TVRSVQKAAASLHELCAARVSE 334
 DB 317 ITKQQTART-YELSSPCLPQGTAEGERVKISDDEIVTVNFKGKVPAG--PQCRYLADK 373
 QY 335 VLQ-----NVRHTEEV---KXVDFAPSVYVDLAGVGLDAEKGSLVVG 378
 DB 374 ILNWDKCNTPPCSPFNGIHQPSLVHTEKESDLYVFSIFDRTOPLG----PLSFTLQ 428
 QY 379 DEFLAKYVCRTE-----TOPSSPSCMDLTVV-VSLLOEFGPRSKVLK 424
 DB 429 ELQDLARVTCNGEHWESVFGISLSLSKPKQWCIDLNFQVSLHTGYDILQELR 488
 QY 425 LTRKIDNVETSMALGAIFHYIDSLN 449
 DB 489 TAKTIANNELGMCGLASPLLESAN 513

RESULT 14
 Q8H7L6 PRELIMINARY; PRT; 489 AA.
 ID Q8H7L6;
 AC Q8H7L6;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Putative apyrase.
 GN OSJNB0014110.10.
 OS Oryza sativa (Japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhacoidae; Oryzaceae; Oryza.
 NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wing R.A., Yu Y., Soderlund C., Kim H.-R., Rambo T., Sacki C.,
 RA Currie J., Colliura K.;
 RT "Rice Genomic Sequence.";
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC126222; AAN65004.1; -;
 DR InterPro: IPR000407; GDAI_CD39_NTPase.
 DR Pfam; PF01150; GDAI_CD39_1.
 DR PROSITE; PS01238; GDAI_CD39_NTPase; 1.
 DR PS01238; GDAI_CD39_NTPase; 1.
 SQ SEQUENCE 489 AA; 52799 MW; 0BF6BF15448D398 CRC64;

Query Match 22.3%; Score 528; DB 10; Length 489;
 Best Local Similarity 33.0%; Pred. No. 7,2e-37;
 Matches 145; Conservative 80; Mismatches 157; Indels 58; Gaps 14;

QY 49 AAPGARWGOAHSPLGTADGHEVYGYGIMPDAAGTGRVAVFOFTRPREPTLTH-ETF 107
 DB 66 AAGGRWGPRAVSGG--DG-STRYAVIFDAGSSGSRVHYCFDGNLDLPIGKEILF 121
 QY 108 KAVKGLSAYADVEKSAQGIRELVDVAKODIPDFWKATPLVKATAGRLPGEKAQK 167
 DB 122 KQKRGSLSYAMDPEAKSLVSLLEBAKVIPLVLRQGTIPRVATAGRLGTEKSEE 181
 QY 168 LIQKVEVFK-ASPVLVGDVCSVINGTDEGVASMTINFLTGSJLTPGSSVGMIDLQ 226
 DB 182 IIQAVVDDLQDSSSPRSGPEWTVLDGQGEAFQWVTINYLGNIGKPKSYHTVGAVDGG 241
 QY 227 GSIQIAFLPRVAGTILQASP-----PGYLTALMFMNRTYKLSYSYLGILMSARLAILGG 281

Db 242 GSVQMAVASEKADGAPPAEGBDSYVKELLLKGTYYLVVHSYLRGLLAARAEILKA 301
Qy 282 VEGQPAKCKELVSPCLSPSPKGEWEHAEVTVYRVSQKAAASLHELCAARYSEVLQNRVH 341
Db 302 GEGNDYRN-----CMLEGHGQRYGDDIFEASGLSSGASYSKRAVAV-----RAL 348
Qy 342 RTEE--VRHV-----DFYAFSYYYD--LAAGVGLIDAEKGGSLV--VGDPE 381
Db 349 KVDEPACTHMKCTFGGVWNGGGGQKQLFVASFFFDRAAEAGFVNPKAPFAKVPKPSDFE 408
Qy 382 IAAKYVCR-----TLETQPS--PFSCMDLY-VSLLLOEFGFPRSKVLKTRKI-- 429
Db 409 EAARVCKLVKNDQAQATYPDVSEENVYCLMDLVYQYTLVVDGFGVDYQDITLVKKVPY 468
Qy 430 --DNVETSWALGAIFHYIDS 447
Db 469 SNSFVEAAWPLGSAIEVASS 488

RESULT 15
Q9SPM7 PRELIMINARY; PRT; 467 AA.
AC Q9SPM7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Apyrase.
OC Dolichos biflorus (Horse gram).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Dolichos.
OX NCBI_TaxID=3840;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99444909; PubMed=10517321;
RA Roberts N.J., Brigham J., Wu B., Murphy J.B., Volpin H.,
RA Phillips D.A., Etzler M.E.;
RT "A Nod factor-binding lectin is a member of a distinct class of
RT apyrases that may be unique to the legumes."
RL Mol. Gen. Genet. 262:261-267 (1999).
DR EMBL; AF156781; AAF00610.1; -;
DR InterPro; IPR000407; GDAL_CD39_NTPase.
DR Pfam; PF01150; GDAL_CD39; 1.
DR PROSITE; PS01238; GDAL_CD39_NTPASE; 1.
SQ SEQUENCE 467 AA; 51164 MW; 8FCC200AA60D7376 CRC64;

Query Match 22.2%; Score 525.5; DB 10; Length 467;
Best Local Similarity 31.7%; Pred. No. 1.1e-36;
Matches 145; Conservative 85; Mismatches 162; Indels 65; Gaps 17;
Qy 37 ATATQAFSITRAAPGARWQQQAHSPGTAADGHEV-----FYGIMFDAGSTGRVHV 90
Db 21 ATATASFSL--HGKFK-----HRKFSSDNNYSIEETINESYAVIFDAGSTGRVHV 73
Qy 91 QFTRPREPTTLTH--ETPKAVKPGLSAYADVDEKSAQCIRELLDVAKODIPDFWKATP 148
Db 74 RFNQ-QLDLRLIGHDLLELFVKTKPGLSAYAPENPEAAESLVPLLEAEAVIPQELHPTP 132
Qy 149 LVLKATAGLRLPGEKAQKLQKVEVP-KASPFVGDGCVSIMNGTDEGVSAWITINFL 207
Db 133 VKVGATAGLRLQLEGDSNRILQVSDMLKRLTKLVEGDAVSVLSGNQEGAYQWVITINYL 192
Qy 208 TGS�KTPGGSSVGMPLDGGGSGTQIAFLPRVEGTQAS--PPG---YLTALRMFNFTKLY 262
Db 193 LGNLGKHSYKTVAVVDLGGGSVQMAVAISEDAAKAPQVDPGVESYITEMFLRGKKYLY 252
Qy 263 SYSYGLGLMSARLAILGVEGPAKDGKELVSPCLSPSKGEWEHAEVTVYRVSQKAAA 322
Db 253 VHSYLRGLLAARAEVL-----KVRSDSENPCILSGFDGTYTYGGVQYKATAPPSSG 304
Qy 323 SLHELCAARYSEVLQNRVHRTVEVKHV-----DFYAFSYYYDLAAGVGLI 367

Db 305 SFSK-QQNVVLEAL--HVNATCSYKDCPTGGIWNQGGGAGENNPFVASFVEVDEAGFV 361
Qy 368 DAKEGGSLV-VGDFEIAAKYVCRTELETQPS--PFSCMDLY-VSLLLOEF 415
Db 362 DPNDANAIVRPYDFEDAAKACST-ELKDKSVFPRVKGDVPIYICLDLVYQVTLVVDGF 420
Qy 416 GFPRSKVLKTRKIDN-----VETSWALGAIFHYIDS 448
Db 421 GIDPQOEITLVROIQYQDSLVLEAAWPLGSAIEAISL 457

Search completed: June 8, 2004, 09:51:33
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: June 8, 2004, 09:48:14 ; Search time 59 Seconds
(Without alignments)
2183.757 Million cell updates/sec

Title: US-09-905-589A-2

Perfect score: 2364
Sequence: 1 MRRISMHGSLRVAKVAVPLG.....ALGATPHYIDSLNRKSPAS 456

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2364	100.0	456	5	AAE19881 Human CD3
2	2364	100.0	484	4	AAE19881 Human CD3
3	2361	99.9	463	5	ABO6124 Human NS
4	2361	99.9	467	5	ABJ04657 Protein o
5	2351	99.5	456	4	AAm93929 Human pol
6	2152	91.0	450	7	ADCI4220 Human enz
7	2135	90.3	462	4	AAU30882 Novel hum
8	2116	89.5	446	5	ABJ04658 Protein o
9	2003	84.7	471	4	AAE19881 Human CD3
10	999	42.3	428	3	AAE19881 Human CD3
11	999	42.3	428	4	AAE19881 Human CD3
12	999	42.3	428	5	AAE19881 Human CD3
13	999	42.3	428	5	AAE19881 Human CD3
14	996	42.1	428	3	AAE19881 Human CD3
15	996	42.1	428	4	AAE19881 Human CD3
16	909	38.5	405	3	AAE19881 Human CD3
17	909	38.5	405	4	AAE19881 Human CD3
18	904	38.2	465	5	AAE19881 Human CD3
19	823	34.8	330	3	AAE19881 Human CD3
20	621	26.3	461	4	AAE19881 Human CD3
21	621	26.3	461	4	AAE19881 Human CD3
22	596	25.2	144	4	AAU30881 Novel hum
23	518.5	21.9	467	2	AAW85687 DBX oligo
24	506	21.4	462	2	AAW85684 NBP46 roo

26	506	21.4	462	5	AAU78818 Dolichos
27	501	21.2	486	2	AAW85685 NBP46 roo
28	501	21.2	486	5	AAU78818 Lotus jap
29	488.5	20.7	472	6	ABP81286 Arabidops
30	486	20.6	457	7	AAE19881 Human CD3
31	477	20.2	407	7	AAE19881 Human CD3
32	463	19.6	496	5	AAU78820 Medicago
33	454	19.2	496	2	AAW85686 NBP46 roo
34	441.5	18.7	455	2	AAE19881 Human CD3
35	414	17.5	105	4	ABP69389 Human pol
36	371	15.7	139	4	AAU30880 Novel hum
37	314.5	13.3	529	5	AAE19881 Human CD3
38	314.5	13.3	529	5	AAU76973 Human CD3
39	304.5	12.9	604	5	ABP64710 Human pro
40	304.5	12.9	604	5	AAE19881 Human CD3
41	304.5	12.9	604	7	AAU09133 Novel hum
42	287.5	12.2	478	3	AAU70914 Human sol
43	287.5	12.2	478	3	AAU70891 Protein e
44	287	12.1	476	3	AAU70912 Human CD3
45	287	12.1	476	3	AAU70889 Protein e

ALIGNMENTS

RESULT 1	AAE19881	Standard, protein; 456 AA.
ID	AAE19881	Standard, protein; 456 AA.
AC	AAE19881	
XX		
DT	18-JUN-2002	(first entry)
XX		
DE	Human CD39L2 protein.	
XX		
XX	Human, CD-39-like protein; CD39L2 protein; therapy; immune deficiency;	
KW	autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;	
KW	rheumatoid arthritis; autoimmune thyroiditis; allergic reaction; asthma;	
KW	insulin dependent diabetes mellitus; periodontal disease; osteoporosis;	
KW	osteoarthritis; wound healing; tissue repair; Alzheimer's disease; ulcer;	
KW	Parkinson's disease; amyotrophic lateral sclerosis; Huntington's disease;	
KW	nervous system disease; nerve injury; ischaemia-reperfusion injury;	
KW	endotoxin lethality; arthritis; nephritis; inflammatory bowel disease;	
KW	Cronh's disease; vitricide; antibacterial; antifungal; neuroprotective;	
KW	dermatological; immunosuppressive; vulnery; noctropic; anticonvulsant;	
KW	antiinflammatory; nephrotropic; gastrointestinal; vasotropic.	
XX		
OS	Homo sapiens.	
XX		
PN	US6350447-B1.	
XX		
PD	26-FEB-2002.	
XX		
PF	29-JAN-1999; 99US-00240639.	
XX		
PR	29-JAN-1999; 99US-00240639.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Chadwick BP, Frieauff A;	
XX		
DR	WPI, 2002-215262/27.	
XX		
DR	N-PDB; AAD1693.	
XX		
PT	An isolated polypeptide with phosphohydrolase activity, designated	
PT	CD39L2, useful to identify other proteins with which binding occurs or	
PT	identify inhibitors and for treatment of, e.g., Alzheimer's, multiple	
PT	sclerosis and osteoporosis.	
XX		
PS	Claim 1, Fig 4; 10pp; English.	
XX		
CC	The present invention relates to novel proteins with phosphohydrolase	
CC	activity, designated CD-39-like (CD39L) proteins and polynucleotides	

CC encoding such proteins, CD39L proteins are useful to treat infectious
 CC diseases caused by viral, bacterial, fungal or other infection that may
 CC be treatable with CD39L. They are useful in the treatment of various
 CC immune deficiencies and disorders, autoimmune disorders such as multiple
 CC sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune
 CC thyroiditis and insulin dependent diabetes mellitus, allergic reactions
 CC and conditions such as asthma and other respiratory problems, periodontal
 CC disease, osteoporosis, osteoarthritis and other tooth repair processes.
 CC They may have utility in compositions used for bone, cartilage, tendon,
 CC ligament and/or nerve tissue growth or regeneration as well as for wound
 CC healing and tissue repair and replacement and in the treatment of burns,
 CC incisions and ulcers. CD39L proteins may also be useful for proliferation
 CC of neural cells and for regeneration of nerve and brain tissue, i.e. for
 CC the treatment of central nervous system diseases such as Alzheimer's
 CC disease, Parkinson's disease, amyotrophic lateral sclerosis, Huntington's
 CC disease, peripheral nervous system diseases peripheral nerve injuries,
 CC peripheral neuropathy and localised neuropathies. They are also used to
 CC treat mechanical and traumatic disorders which involve degeneration,
 CC death or trauma to neural cells or nerve tissue. CD39L proteins of the
 CC invention are also useful to promote better or faster closure of non-
 CC healing wounds, including pressure ulcers, ulcers associated with
 CC vascular insufficiency and surgical and traumatic wounds. They also
 CC exhibit anti-inflammatory activity and may be used to treat inflammatory
 CC conditions including chronic or acute conditions), including ischaemia-
 CC reperfusion injury, endotoxin lethality, arthritis, nephritis, cytokine
 CC or chemokine-induced lung injury, inflammatory bowel disease or Crohn's
 CC disease. The present sequence is human CD39L2 protein
 XX
 XX
 SQ Sequence 456 AA;

Query Match 100.0%; Score 2364; DB 5; Length 456;
 Best Local Similarity 100.0%; Pred. No. 5.4e-216;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFSTRAAPGARGWQQA 60
 DB 1 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFSTRAAPGARGWQQA 60
 QY 61 SPLGTADGHEVYGINFDAGSGTGVHVQFTRPPRETPTLTHETFKAVKPGLSAYADD 120
 DB 61 SPLGTADGHEVYGINFDAGSGTGVHVQFTRPPRETPTLTHETFKAVKPGLSAYADD 120
 QY 121 VESAGIIRLLDVAKQDIPDFPWKATPLVLKATAGLLPGEKAKLQKVEFKASP 180
 DB 121 VESAGIIRLLDVAKQDIPDFPWKATPLVLKATAGLLPGEKAKLQKVEFKASP 180
 QY 181 FLVGDDCVSIMNGTDEGSAWITINFLTGLSKTPGGSSVGMLDLGGGSTQIAFLPRVGT 240
 DB 181 FLVGDDCVSIMNGTDEGSAWITINFLTGLSKTPGGSSVGMLDLGGGSTQIAFLPRVGT 240
 QY 241 LQASPPGYLTALRMFRTVKLYSYVLGLGLMSARLAILGGVEGPAKDGKELVSPCLSP 300
 DB 241 LQASPPGYLTALRMFRTVKLYSYVLGLGLMSARLAILGGVEGPAKDGKELVSPCLSP 300
 QY 301 SFGGEWEAEVTVYRVSQKAAASLHELCAARVSEVLQNRVHRTVEEVKHYVDYFAFYDDL 360
 DB 301 SFGGEWEAEVTVYRVSQKAAASLHELCAARVSEVLQNRVHRTVEEVKHYVDYFAFYDDL 360
 QY 361 AAGVGLIDAKGSLVVGDFEIAAKVCTLETPQSPSPFSCMDLTYVSLILQEGFPFRS 420
 DB 361 AAGVGLIDAKGSLVVGDFEIAAKVCTLETPQSPSPFSCMDLTYVSLILQEGFPFRS 420
 QY 421 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456
 DB 421 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456

RESULT 2
 AAB72241
 ID AAB72241 standard; protein; 484 AA.
 XX
 AC AAB72241;
 XX

DT 14-MAY-2001 (first entry)
 XX Human CD39 like protein CD39-L2 amino acid sequence.
 DE Human CD39-like protein; apyrase; NDPase; platelet function inhibitor;
 XX myocardial infarction; cerebral ischaemia; angina; arterial thrombosis;
 KW cerebral artery thrombosis; platelet aggregation; inflammation;
 KW apoptosis; autoimmune disorder; neurological disorder;
 KW Alzheimer's disease; Parkinson's disease; cancer; CD39-L2.
 XX Homo sapiens.
 OS WO200110205-A1.
 XX 15-FEB-2001.
 XX 09-AUG-2000; 2000WO-US021790.
 XX 09-AUG-1999; 99US-00370265.
 PR 11-JAN-2000; 2000US-00481238.
 PR 25-APR-2000; 2000US-00557800.
 PR 26-MAY-2000; 2000US-00583231.
 PR 30-JUN-2000; 2000US-00608285.
 XX (HYSE-) HYSEQ INC.
 PA Ford J, Mulero JJ, Yeung G;
 XX WPI; 2001-147489/15.
 XX N-PSDB; AAF63386.
 XX Polynucleotides encoding human CD39-like polypeptides, with apyrase
 PT and/or NDPase activity, which are useful in the treatment of pathological
 PT conditions caused by thrombosis (e.g. myocardial infarction) and
 PT inflammatory disorders.
 PS Claim 39; Page 162-164; 203pp; English.
 XX This invention relates to polynucleotides encoding human CD39-like
 CC polypeptides with apyrase and/or NDPase activity. The polypeptides having
 CC ATPase, including NDPase, activity are useful for inhibiting platelet
 CC function and can therefore be used in the prophylaxis or treatment of
 CC pathological conditions caused by or involving thrombosis or excessive
 CC coagulation or excessive platelet aggregation, such as myocardial
 CC infarction, cerebral ischaemia, angina, arterial thrombosis, cerebral
 CC artery thrombosis or intracardiac thrombosis, and conditions associated
 CC with venous thrombosis. CD39-L4 and CD39-L2 polypeptides are useful in
 CC modulating disease states (including platelet aggregation, inflammation
 CC and apoptosis) associated with ADP or other purinergic signalling by
 CC reducing the levels of NDPs. The polypeptides are also useful for
 CC prophylaxis or treatment of inflammation related disorders, such as
 CC disorders involving sepsis or systemic inflammatory response syndrome or
 CC SIRS (and associated conditions such as fever, tachycardia, tachypnea,
 CC cytokine overstimulation); autoimmune disorders such as thrombosis,
 CC atherosclerosis, acute pancreatitis, dermatitis, including psoriasis,
 CC cirrhosis, reperfusion injury, asthma, multiple sclerosis, arthritis;
 CC neurological disorders including neurodegenerative diseases, epilepsy,
 CC depression, Alzheimer's disease, Parkinson's disease, Huntington's
 CC disease, and amyotrophic lateral sclerosis; and cancer. The present
 CC sequence represents human CD39 like protein CD39-L2
 XX Sequence 484 AA;

Query Match 100.0%; Score 2364; DB 4; Length 484;
 Best Local Similarity 100.0%; Pred. No. 5.9e-216;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFSTRAAPGARGWQQA 60
 DB 29 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFSTRAAPGARGWQQA 88
 QY 61 SPLGTADGHEVYGINFDAGSGTGVHVQFTRPPRETPTLTHETFKAVKPGLSAYADD 120

Db 89 SPLGTADGHEVFGYIMFDAGSTGTRVHVFQFTRPREPTLTTHETFKAVKPGLSAYADD 148
QY 121 VESKAGQIRELLDVAKODIPDFPKATPLVLTAKATGRLPGKAKGLQKVEVKASP 180
Db 149 VESKAGQIRELLDVAKODIPDFPKATPLVLTAKATGRLPGKAKGLQKVEVKASP 208
QY 181 FLVGGDDCVSINMGTEBGSAMITINFLTSGSLKTPGSSVGMGLDGGSTQIAFLPREVGT 240
Db 209 FLVGGDDCVSINMGTEBGSAMITINFLTSGSLKTPGSSVGMGLDGGSTQIAFLPREVGT 268
QY 241 LQASPPGYTLALRMFNRTYKLSYSYGLGMSARLAILGVGQAPAKGKELVSPCLSP 300
Db 269 LQASPPGYTLALRMFNRTYKLSYSYGLGMSARLAILGVGQAPAKGKELVSPCLSP 328
QY 301 SFKGEWEHAETVTVSGOKAASLHELCAARVSEVLQNRVHRTBEVKHVPFAFSYYVDL 360
Db 329 SFKGEWEHAETVTVSGOKAASLHELCAARVSEVLQNRVHRTBEVKHVPFAFSYYVDL 388
QY 361 AAGVGLIDAEKGSLLVGDPEIARVYCRITLETQPOSSPSCMDLTVYSLLQEFGEFPRS 420
Db 389 AAGVGLIDAEKGSLLVGDPEIARVYCRITLETQPOSSPSCMDLTVYSLLQEFGEFPRS 448
QY 421 KVLKLTTRKIDNVTSMALGALFHYIDSLNRQKSPAS 456
Db 449 KVLKLTTRKIDNVTSMALGALFHYIDSLNRQKSPAS 484

RESULT 3
ABR06124
ID ABR06124 standard; protein; 463 AA.
AC ABR06124;
DT 10-MAY-2002 (first entry)
DE Human NS protein sequence SEQ ID NO:216.
XX Human, cytosol; osteopathic; gynaecological; neuroprotective;
KM antitubercular; antitubercular; antiparasitic; ophthalmological; anti-HIV;
KM vasoregulatory; antitubercular; antitubercular; dermatological;
KM anorectic; muscular; antitubercular; cardiovascular; anticoagulant;
KM antifibrinolytic; hypotensive; antitubercular; immunomodulator; cardiac;
KM anticonvulsant; antidiabetic; tranquilizer; antidepressant; neuroleptic;
KM gastroenteric; antitubercular; antitubercular; cerebroprotective; nootropic;
KM contraceptive; vaccine; gene therapy; cancer; osteoporosis; psoriasis;
KM endometriosis; degenerative disease; multiple sclerosis; psoriasis;
KM rheumatoid arthritis; cataract; restenosis; atherosclerosis; glaucoma;
KM inflammation; skin disorder; obesity; muscular dystrophy; AIDS;
KM infertility; cardiovascular disease; coagulation disease; hypertension;
KM ischaemia; asthma; immune disease; epilepsy; angina; neurodegeneration;
KM diabetes; anxiety; depression; schizophrenia; viral disease; stroke;
KM gastric ulcer; Alzheimer's disease.
XX Homo sapiens.
OS WO200206315-A2.
FN 24-JAN-2002.
PD 17-JUN-2001; 2001WO-IL000653.
XX 18-JUN-2000; 2000IL-00137345.
PR 15-DEC-2000; 2000IL-00140354.
XX (COMP-) COMPUGEN LTD.
PA Mintz L, Freilich S, Bernstein J;
PI WPI; 2002-155037/20.
XX N-PSDB; ABL39778.
XX One hundred and twenty eight novel nucleic acid sequences, useful for
PT treating and diagnosing e.g. cancer, asthma and Alzheimer's.

XX Claim 6, Page 251-253; 290pp; English.
PS ABL39691 to ABL39818 represent novel human nucleic acid sequences
XX encoding the proteins given in ABR06037 to ABR06144. The novel sequences
CC (NS) can have cytosol, osteopathic, gynaecological, neuroprotective,
CC antitubercular, antitubercular, antiparasitic, ophthalmological, vaccine,
CC vasoregulatory, antitubercular, antitubercular, dermatological,
CC anorectic, muscular, antitubercular, hypotensive, cardiovascular, cardiac,
CC anticonvulsant, antifibrinolytic, hypotensive, antitubercular, cardiac,
CC immunomodulator, anticonvulsant, antidiabetic, tranquilizer, antitubercular,
CC antidepressant, gastroenteric, antitubercular, cerebroprotective,
CC nootropic and contraceptive activities. The NS can be used in vaccines,
CC gene therapy and antisense therapy. Nucleic acids, expression vectors and
CC antibodies from the present invention can be used for treating and
CC diagnosing e.g. cancer, osteoporosis, endometriosis, degenerative
CC diseases, dystonia, multiple sclerosis, rheumatoid arthritis, psoriasis,
CC cataracts, restenosis, atherosclerosis, inflammation, skin disorders,
CC glaucoma, obesity, muscular dystrophy, AIDS, infertility, cardiovascular
CC disease, coagulation disease, ischaemia, hypertension, asthma, immune
CC depression, schizophrenia, viral disease, gastric ulcers, stroke,
CC Alzheimer's disease and as a contraceptive
XX

SEQ Sequence 463 AA:
SQ Query Match 99.9%; Score 2361; DB 5; Length 463;
Best Local Similarity 99.8%; Pred. No. 1.1e-215;
Matches 455; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKISHGSLRVAKAYPLGLCVGFIVVAYIKMRAATATQAFSITRAPCARMGQAH 60
Db 8 MKKISHGSLRVAKAYPLGLCVGFIVVAYIKMRAATATQAFSITRAPCARMGQAH 67
QY 61 SPLGTADGHEVFGYIMFDAGSTGTRVHVFQFTRPREPTLTTHETFKAVKPGLSAYADD 120
Db 68 SPLGTADGHEVFGYIMFDAGSTGTRVHVFQFTRPREPTLTTHETFKAVKPGLSAYADD 127
QY 121 VESKAGQIRELLDVAKODIPDFPKATPLVLTAKATGRLPGKAKGLQKVEVKASP 180
Db 128 VESKAGQIRELLDVAKODIPDFPKATPLVLTAKATGRLPGKAKGLQKVEVKASP 187
QY 181 FLVGGDDCVSINMGTEBGSAMITINFLTSGSLKTPGSSVGMGLDGGSTQIAFLPREVGT 240
Db 188 FLVGGDDCVSINMGTEBGSAMITINFLTSGSLKTPGSSVGMGLDGGSTQIAFLPREVGT 247
QY 241 LQASPPGYTLALRMFNRTYKLSYSYGLGMSARLAILGVGQAPAKGKELVSPCLSP 300
Db 248 LQASPPGYTLALRMFNRTYKLSYSYGLGMSARLAILGVGQAPAKGKELVSPCLSP 307
QY 301 SFKGEWEHAETVTVSGOKAASLHELCAARVSEVLQNRVHRTBEVKHVPFAFSYYVDL 360
Db 308 SFKGEWEHAETVTVSGOKAASLHELCAARVSEVLQNRVHRTBEVKHVPFAFSYYVDL 367
QY 361 AAGVGLIDAEKGSLLVGDPEIARVYCRITLETQPOSSPSCMDLTVYSLLQEFGEFPRS 420
Db 368 AAGVGLIDAEKGSLLVGDPEIARVYCRITLETQPOSSPSCMDLTVYSLLQEFGEFPRS 427
QY 421 KVLKLTTRKIDNVTSMALGALFHYIDSLNRQKSPAS 456
Db 428 KVLKLTTRKIDNVTSMALGALFHYIDSLNRQKSPAS 463

RESULT 4
ABJ04657
ID ABJ04657 standard; protein; 467 AA.
AC ABJ04657;
DT 11-OCT-2002 (first entry)
XX Protein of NOVX 15a SEQ ID NO 36.
XX

KW Cytostatic; antidiabetic; anorectic; metabolic; nootropic; antilipemic; neuroprotective; antiparkinsonian; anticonvulsant; cerebroprotective; tranquiliser; neuroleptic; antidiabetic; antitumor; antiinflammatory; anti-HIV; antiallergic; antirheumatic; antiarthritic; NOVX; diabetes; metabolic disorder; obesity; infectious disease; Alzheimer's disease; anorexia; neurodegenerative disorder; Parkinson's disorder; obesity; immune disorder; haematopoietic disorder; dyslipidaemia; chronic disease; metabolic syndrome X; wasting disorder; cancer; neurological disorder; epilepsy; stroke; mental disorder; schizophrenic disorders; goiter; vesicular transport; cystic fibrosis; gastrointestinal disorder; diabetes mellitus; ulcerative colitis; AIDS; allergic reaction; multiple sclerosis; rheumatoid arthritis; transgenic animal; gene therapy.

XX Unidentified.

XX WO200246409-A2.

XX 13-JUN-2002.

XX 06-DEC-2001; 2001WO-US046586.

XX 06-DEC-2000; 2000US-0251660P.

PR 12-DEC-2000; 2000US-0255029P.

PR 08-JAN-2001; 2001US-0260326P.

PR 24-JAN-2001; 2001US-0263800P.

PR 20-FEB-2001; 2001US-0269942P.

PR 24-APR-2001; 2001US-0286183P.

PR 20-AUG-2001; 2001US-0313627P.

PR 12-SEP-2001; 2001US-0318712P.

XX (CURA-) CURAGEN CORP.

XX Guo X, Li L, Patturajan M, Shimkets RA, Casman SJ, Malyankar UM;

XX Tchernev VT, Vernet CAM, Spytek KA, Stenoy SG, Alsbrook JP;

XX Edinger S, Peyman JA, Stone DJ, Ellerman K, Gangoli EA, Boldog FI;

XX Colman SD, Eisen AJ, Liu X, Padigaru M, Spaderna SK, Zerhusen BD;

XX WPI: 2002-547774/58.

XX N-PSDB; ABT05470.

XX Novel isolated polypeptide, designated NOVX, useful for treating or

XX preventing cancer, diabetes, obesity, dyslipidemia, anorexia, and

XX metabolic, neurodegenerative, immune and hematopoietic disorders.

XX Claim 1; Page 140; 421pp; English.

XX The invention relates to an isolated polypeptide, designated NOVX,

XX comprising a sequence fully defined in the specification. The isolated

XX protein, its encoding polynucleotide or an antibody created from the

XX protein is useful in the manufacture of a medicament for treating a

XX syndrome associated with a human disease, preferably a NOVX-associated

XX disorder, or for treating or preventing a NOVX-associated disorder in a

XX subject, preferably human. The isolated protein, its encoding

XX polynucleotide or an antibody created from the protein are also useful

XX for treating or preventing metabolic disorders, diabetes, obesity,

XX infectious disease, anorexia, neurodegenerative disorder, Alzheimer's

XX disease, Parkinson's disorder, immune disorders, haematopoietic

XX disorders, and various dyslipidaemias, metabolic disturbances associated

XX with obesity, the metabolic syndrome X, wasting disorders associated with

XX chronic diseases, and cancer. The isolated protein, its encoding

XX polynucleotide or an antibody created from the protein are useful for

XX treating or preventing neurological disorders such as epilepsy, stroke,

XX mental disorders including mood, anxiety, schizophrenic disorders,

XX disorders of vesicular transport such as cystic fibrosis, diabetes

XX mellitus, goiter, gastrointestinal disorders including ulcerative

XX colitis, other conditions associated with abnormal vesicle trafficking

XX including AIDS, allergic reactions, multiple sclerosis and rheumatoid

XX arthritis. A cell comprising the vector of the invention is useful for

XX producing non-human transgenic animals. The polynucleotide of the

XX invention can be used to treat disorders by gene therapy. This sequence

XX represents one of the isolated NOVX proteins of the invention

SQ Sequence 467 AA;
Query Match 99.9%; Score 2361; DB 5; Length 467;
Best Local Similarity 99.8%; Pred. No. 1.1e-215;
Matches 455; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MRKISNHSGLRVAKVAYPLGLCVGVYIYVAYIKWHRATATQAFPSITRAAPGARWGQQA 60
Db 12 MRKISNHSGLRVAKVAYPLGLCVGVYIYVAYIKWHRATATQAFPSITRAAPGARWGQQA 71
Qy 61 SPLGTAADGHEVFGIMFDAGSTGTRVHVQFTRPREPTLTHETPKAVKPCLSAYADD 120
Db 72 SPLGTAADGHEVFGIMFDAGSTGTRVHVQFTRPREPTLTHETPKAVKPCLSAYADD 131
Qy 121 VEKSAQGIREDLDVAKODIPDFWKATPLVLKATAGLRLPGKAKQKLLQKVKEVFKASP 180
Db 132 VEKSAQGIREDLDVAKODIPDFWKATPLVLKATAGLRLPGKAKQKLLQKVKEVFKASP 191
Qy 181 FLYGDDCVSIMGTDGVSAMITINLTGSLKTPGSSVGMGLDGGSTOIAFLPRVEGT 240
Db 192 FLYGDDCVSIMGTDGVSAMITINLTGSLKTPGSSVGMGLDGGSTOIAFLPRVEGT 251
Qy 241 LQASPPGYLTALRMENRTYKLYSYSLGLGMSARLAILGGVGGQPAKQKELVSPCLSP 300
Db 252 LQASPPGYLTALRMENRTYKLYSYSLGLGMSARLAILGGVGGQPAKQKELVSPCLSP 311
Qy 301 SFKGWEHAETVTVRSQKAAASLHELCAARSEVQLNVRHRTTEVKHVDYFAFSYYDL 360
Db 312 SFKGWEHAETVTVRSQKAAASLHELCAARSEVQLNVRHRTTEVKHVDYFAFSYYDL 371
Qy 361 AAGVGLLDAEKGSLVVGDFEIAAKVVCRTLETQPOSSPSCMDLTVVSLLLQEFQFPFRS 420
Db 372 AAGVGLLDAEKGSLVVGDFEIAAKVVCRTLETQPOSSPSCMDLTVVSLLLQEFQFPFRS 431
Qy 421 KVLKLTPIKIDNVTETSWALGAIFHYIISLNQKSPAS 456
Db 432 KVLKLTPIKIDNVTETSWALGAIFHYIISLNQKSPAS 467
RESULT 5
AAM93929 standard; protein; 456 AA.
ID AAM93929 standard; protein; 456 AA.
AC AAM93929;
XX AAM93929;
DT 06-NOV-2001 (first entry)
XX Human polypeptide, SEQ ID NO: 4100.
DE Human; full length cDNA; cDNA synthesis; oligo-capping.
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX Homo sapiens.
OS Homo sapiens.
XX EP1130094-A2.
XX 05-SEP-2001.
PD 05-SEP-2001.
PF 07-JUL-2000; 2000EP-00114089.
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX (HELI-) HELIX RES INST.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.
DR N-PSDB; AAK94892.
XX 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.

XX Claim 8; SEQ ID NO 4100; 1380bp + Sequence Listing; English.
PS
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5' - and 3' - ends of the cDNA molecules have
CC been determined. Primers for synthesizing the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is a polypeptide encoded by a full length
CC human cDNA of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in CD-ROM
CC format directly from EPO
CC
XX
SQ Sequence 456 AA;
Query Match 99.5%; Score 2351; DB 4; Length 456;
Best Local Similarity 99.3%; Pred. No. 9.3e-215;
Matches 453; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MRKISNHSGLRVAKAYAPLGLCVGFIVAYIKWHRATATQAFPSITRAAPGARWGQAH 60
DB 1 MRKISNHSGLRVAKAYAPLGLCVGFIVAYIKWHRATATQAFPSITRAAPGARWGQAH 60
QY 61 SPLGTADGHEVFYIGIMFDAGSTGRVAVFOFTRPPRETPTLTHETFRKAVKPGISAYAD 120
DB 61 SPLGTADGHEVFYIGIMFDAGSTGRVAVFOFTRPPRETPTLTHETFRKAVKPGISAYAD 120
QY 121 VEKSAOGIREHLDVAKODIPEDFKAPPLVAKATAGRLILPGEKAKLLOKVEVFASFP 180
DB 121 VEKSAOGIREHLDVAKODIPEDFKAPPLVAKATAGRLILPGEKAKLLOKVEVFASFP 180
QY 121 VEKSAOGIREHLDVAKODIPEDFKAPPLVAKATAGRLILPGEKAKLLOKVEVFASFP 180
DB 181 FLVGGDCVSIWNGTDEGVSAWITINFLTSLKTPGSSVGMLDJGGSTQIAFLPRVEGT 240
DB 181 FLVGGDCVSIWNGTDEGVSAWITINFLTSLKTPGSSVGMLDJGGSTQIAFLPRVEGT 240
QY 241 LQASPGVLTALRPNRTYKYSYSLGIGMSARLITIGVEGPKKDGELVSPCLSP 300
DB 241 LQASPGVLTALRPNRTYKYSYSLGIGMSARLITIGVEGPKKDGELVSPCLSP 300
QY 301 SFKGEWEAEVTVYVSGQKAAASHLCAARVSEVLQNRVHRTVEVXKVDVFAYSYTDL 360
DB 301 SFKGEWEAEVTVYVSGQKAAASHLCAARVSEVLQNRVHRTVEVXKVDVFAYSYTDL 360
QY 361 AAGVGLIDAEKGGSLVGDPEIANKYVCTRETETPOSSPFSQMDLTVYSLILQEPFPRS 420
DB 361 AAGVGLIDAEKGGSLVGDPEIANKYVCTRETETPOSSPFSQMDLTVYSLILQEPFPRS 420
QY 421 KYLKLTRKIDIVETSMALGFHYIDSINROKSPAS 456
DB 421 KYLKLTRKIDIVETSMALGFHYIDSINROKSPAS 456
RESULT 6
ADCL4220
ID ADCL4220 standard; protein, 450 AA.
AC ADCL4220;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human enzyme ENZM-26.
XX
KM enzyme; human; ENZM; cytosolic; antiarteriosclerotic; antidiabetic;
KM anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;
KM antiallergic; antiinflammatory; thymimetic; gene therapy;
KM cell proliferative disorder; endocrine disorder; neurological disorder;
KM immune system disorder; inflammatory disorder; developmental disorder;
KM reproductive disorder; vesicle-trafficking disorder; infection.
XX
OS Homo sapiens.

XX
PN WO2003042357-A2.
XX
XX
PD 22-MAY-2003.
XX
XX
PF 26-SEP-2002; 2002WO-US011096.
XX
XX
PR 26-SEP-2001; 2001US-0326388P.
XX
PR 12-OCT-2001; 2001US-0328979P.
XX
PR 19-OCT-2001; 2001US-0346034P.
XX
PR 26-OCT-2001; 2001US-0348284P.
XX
PR 08-NOV-2001; 2001US-0338048P.
XX
PR 16-NOV-2001; 2001US-0332340P.
XX
PR 14-DEC-2001; 2001US-0340357P.
XX
PR 29-MAR-2002; 2002US-0368722P.
XX
PR 29-MAR-2002; 2002US-0368799P.
XX
PR 17-MAY-2002; 2002US-0381588P.
XX
PR 07-JUN-2002; 2002US-0387119P.
XX
PR 21-JUN-2002; 2002US-0390662P.
XX
XX
PA (INCYTE) INCYTE GENOMICS INC.
XX
PI Yang J, Lu DM, Yue H, Elliott VS, Warren BA, Duggan BM;
PI Forsythe JD, Lee EA, Hafalia AUA, Ramkumar J, Chawla NK, Baughn MR;
PI Becha SD, Gorvad AE, Tran UK, Li JX, Yao MG, Ison CH, Griffin JA;
PI Lee SY, Chang H, Emerling BM, Tang YT, Lal PG, Kable AE;
PI Marquis JP, Jiang X, Jackson AA, Zebadjadian Y, Satriakkar A,
PI Wilson AD, Jin P, Richardson TW, Bhatia U, Burrill JD, Lee S;
PI Blake JJ, Ho A, Zheng W, Gao J;
XX
XX
DR WPI; 2003-449567/42.
DR
DR N-PsDB; ADCL4273.
XX
XX
PT New human enzymes (ENZM), useful for diagnosing, treating and preventing
PT diseases or conditions associated with the aberrant ENZM expression e.g.
PT cancer, diabetes, epilepsy, or infections.
XX
XX
PS Claim 1; SEQ ID NO 26; 416pp; English.
XX
XX
CC The invention relates to a novel isolated human enzyme (ENZM)
CC polypeptide. A polypeptide of the invention has cytosolic,
CC antidiabetic, anticonvulsant, nootropic,
CC antidiabetic, anticonvulsant, neuroprotective, anti-HIV, antiallergic,
CC neuroprotective, cerebroprotective, anti-HIV, antiallergic,
CC antiinflammatory, and thymimetic activity. A polynucleotide encoding a
CC polypeptide of the invention may have a use in gene therapy. The
CC polypeptide and polynucleotides are useful in diagnosing, treating and
CC preventing diseases or conditions associated with the decreased
CC expression or overexpression of ENZM, such as cell proliferative (e.g.
CC cancer, atherosclerosis), endocrine (e.g. diabetes), neurological (e.g.
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC allergies), developmental (e.g. Hypothyroidism, Cushing's syndrome),
CC reproductive and vesicle-trafficking disorders, or infections. These are
CC also useful in assessing the effects of exogenous compounds on the
CC expression of nucleic acid and amino acid sequences of ENZM. The ENZM or
CC its fragments are useful in screening compounds for effectiveness as
CC agonist or antagonist of the polypeptides, or in altering the expression
CC of the target polynucleotide and compounds that specifically bind to or
CC modulate the activity of the polypeptide. The microarray is useful in
CC monitoring or measuring protein-protein interactions, drug-target
CC interactions, and gene expression profiles. The sequences shown in
CC ADCL4195-ADCL4247 represent ENZM proteins of the invention.
XX
XX
SQ Sequence 450 AA;
Query Match 91.0%; Score 2152; DB 7; Length 450;
Best Local Similarity 92.3%; Pred. No. 8.3e-196;
Matches 421; Conservative 1; Mismatches 0; Indels 34; Gaps 1;
QY 1 MRKISNHSGLRVAKAYAPLGLCVGFIVAYIKWHRATATQAFPSITRAAPGARWGQAH 60
DB 29 MRKISNHSGLRVAKAYAPLGLCVGFIVAYIKWHRATATQAFPSITRAAPGARWGQAH 88
QY 61 SPLGTADGHEVFYIGIMFDAGSTGRVAVFOFTRPPRETPTLTHETFRKAVKPGISAYAD 120

Db 89 SPLGTAADGHEVFYGMFDAGSTGTRVHVFPQTRPRETPTLTHTETFKALPKGLSAYADD 148
Qy 121 VEKSAQGIKRELLDVAQKQDIPDFWKATPLVLKATAGLRLPGEKAOKLQKVEVFKASP 180
Db 149 VEKSAQGIKRELLDVAQKQDIPDFWKATPLVLKATAGLRLPGEKAOKLQKVEVFKASP 208
Qy 181 FLVGDDCVSINMGTDGVSAMITINFLTGSLKTPGGSSVGMGLDGGSTQIAFLPRVEGT 240
Db 209 FLVGDDCVSINMGTDGVSAMITINFLTGSLKTPGGSSVGMGLDGGSTQIAFLPRVEGT 268
Qy 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGVEGQPAKDGKELVSPCLSP 300
Db 269 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGVEGQPAKDGKELVSPCLSP 314
Qy 301 SFKGEWEHAEVTVYRSGOKAAASLHELCAARVSEVLQNRVHRTVEVKHVDYFAFSYYDL 360
Db 315 -----AASLHELCAARVSEVLQNRVHRTVEVKHVDYFAFSYYDL 354
Qy 361 AAGVGLIDAEKGGSLVVGDFEIAAKYVCRITLETQPSPPSCMDLTVVSLLLQEFPPRS 420
Db 355 AAGVGLIDAEKGGSLVVGDFEIAAKYVCRITLETQPSPPSCMDLTVVSLLLQEFPPRS 414
Qy 421 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456
Db 415 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 450

RESULT 7
AAU30882
ID AAU30882 standard; protein; 462 AA.
XX
AC AAU30882;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #1373.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
XX vaccination, testing and therapy.
PS Claim 20; Page 365; 765pp; English.
XX

The invention relates to novel human secreted polypeptides. The polypeptides and antibodies to the polypeptides are useful for determining the presence of or predisposition to a disease associated with altered levels of polypeptide. The polypeptides are also useful for identifying agents (agonists and antagonists) that bind to them. Cells expressing the proteins are useful for identifying a therapeutic agent for use in treatment of a pathology related to aberrant expression or physiological interactions of the polypeptide. Vectors comprising the nucleic acids encoding the polypeptides and cells genetically engineered

to express them are also useful for producing the proteins. The proteins are useful in genetic vaccination, testing and therapy, and can be used as nutritional supplements. They may be used to increase stem cell proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon and/or nerve tissue growth or regeneration; immune suppression and/or stimulation; as anti-inflammatory agents; and in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid sequences of novel human secreted proteins of the invention

XX
SQ Sequence 462 AA;
Query Match 90.3%; Score 2135; DB 4; Length 462;
Best Local Similarity 92.1%; Pred. No. 3, 6e-194;
Matches 422; Conservative 2; Mismatches 8; Indels 26; Gaps 3;
Qy 1 MRKISNHGSLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFFSITRAAPCARWQQAH 60
Db 29 MRKISNHGSLRVAKVAYPLGLCVGVFIYVAYIKWHRANATQAFFSITRAAPCARWQQAH 88
Qy 61 SPLGTAADGHEVFYGMFDAGSTGTRVHVFPQTRPRETPTLTHTETFKAVKPLSAYADD 120
Db 89 SPLGTAADGHEVFYGMFDAGSTGTRVHVFPQ-----YADD 124
Qy 121 VEKSAQGIKRELLDVAQKQDIPDFWKATPLVLKATAGLRLPGEKAOKLQKVEVFKASP 180
Db 125 VEKSAQGIKRELLDVAQKQDIPDFWKATPLVLKATAGLRLPGEKAOKLQKVEVFKASP 184
Qy 181 FLVGDDCVSINMGTDGVSAMITINFLTGSLKTPGGSSVGMGLDGGSTQIAFLPRVEGT 240
Db 185 FLVGDDCVSINMGTDGVSAMITINFLTGSLKTPRRSVGMGLDGGSTQIVFLTHVEGT 244
Qy 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGVEGQPAKDGKELVSPCLSP 300
Db 245 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGVEGQPAKDGKELVSPCLSP 304
Qy 301 SFKGEWEHAEVTVYRSGOKAAASLHELCAARVSEVLQNRVHRTVEVKHVDYFAFSYYDL 360
Db 305 SFKGEWEHAEVTVYRSGOKAAASLHELCAARVSEVLQNRVHRTVEVKHVDYFAFSYYDL 364
Qy 361 AAGVGLIDAEKGGSLVVGDFEIAAKYVCRITLETQPSPPSCMDLTVVSLLLQEFQFP 418
Db 365 AAGVGLIDAEKGGSLVVGDFEIAAKYVCRITLETQPSPPSCMDLTVVSLLLQEFQFP 424
Qy 419 RSKVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456
Db 425 RSKVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 462

RESULT 8
ABU04658
ID ABU04658 standard; protein; 446 AA.
XX
AC ABU04658;
XX
DT 11-OCT-2002 (first entry)
XX
DE Protein of NOVX 15b SEQ ID No 38.
XX

Cytostatic; antidiabetic; anorectic; metabolic; nootropic; antilipaeamic; neuroprotective; antiparkinsonian; anticonvulsant; cerebroprotective; tranquiliser; neuroleptic; antidiabetic; antiulcer; antiinflammatory; anti-HIV; antiallergic; antirheumatic; antiarthritic; NOVX; diabetes; metabolic disorder; obesity; infectious disease; Alzheimer's disease; anorexia; neurodegenerative disorder; Parkinson's disorder; obesity; immune disorder; haematopoietic disorder; dyslipidaemia; chronic disease; metabolic syndrome X; wasting disorder; cancer; neurological disorder; epilepsy; stroke; mental disorder; schizophrenic disorders; goiter; vesicular transport; cystic fibrosis; gastrointestinal disorder; diabetes mellitus; ulcerative colitis; AIDS; allergic reaction; multiple sclerosis; rheumatoid arthritis; transgenic animal; gene therapy.
XX
OS Unidentified.

XX MO200246409-A2.
 XX 13-JUN-2002.
 XX 06-DEC-2001; 2001WO-US046586.
 XX 06-DEC-2000; 2000US-0251660P.
 XX 12-DEC-2000; 2000US-0255029P.
 XX 08-JAN-2001; 2001US-0260326P.
 XX 24-JAN-2001; 2001US-0263800P.
 XX 20-FEB-2001; 2001US-0269942P.
 XX 24-APR-2001; 2001US-0286183P.
 XX 20-AUG-2001; 2001US-0313627P.
 XX 12-SEP-2001; 2001US-0318712P.
 XX (CURA-) CURAGEN CORP.
 XX Guo X, Li L, Paturajan M, Shinkets RA, Casman SJ, Malyankar UM,
 XX Tchernov ST, Vernet CAM, Spytek KA, Shenoy SG, Alsbrook JP,
 XX Edinger S, Peyman JA, Stone DJ, Ellerman K, Gangoli EA, Boldog FL,
 XX Colman SD, Eisen AJ, Liu X, Padigaru M, Spaderna SK, Zernhusen BD;
 XX WPI: 2002-547774/58.
 XX N-PSDB; AB705471.
 XX Novel isolated polypeptide, designated NOVX, useful for treating or
 XX preventing cancer, diabetes, obesity, dyslipidemia, anorexia, and
 XX metabolic, neurodegenerative, immune and hematopoietic disorders.
 XX Claim 1; Page 141; 421pp; English.
 XX The invention relates to an isolated polypeptide, designated NOVX,
 XX comprising a sequence fully defined in the specification. The isolated
 XX protein, its encoding polynucleotide or an antibody created from the
 XX protein, is useful in the manufacture of a medicament for treating a
 XX syndrome associated with a human disease, preferably a NOVX-associated
 XX disorder, or for treating or preventing a NOVX-associated disorder in a
 XX subject, preferably human. The isolated protein, its encoding
 XX polynucleotide or an antibody created from the protein are also useful
 XX for treating or preventing metabolic disorders, diabetes, obesity,
 XX infectious disease, anorexia, neurodegenerative disorder, Alzheimer's
 XX disease, Parkinson's disorder, immune disorders, haematopoietic
 XX disorders, and various dyslipidemias, metabolic disturbances associated
 XX with obesity, the metabolic syndrome X, wasting disorders associated with
 XX chronic diseases, and cancer. The isolated protein, its encoding
 XX polynucleotide or an antibody created from the protein are useful for
 XX treating or preventing neurological disorders such as epilepsy, stroke,
 XX mental disorders including mood, anxiety, schizophrenic disorders,
 XX disorders of vesicular transport such as cystic fibrosis, diabetes
 XX mellitus, goiter, gastrointestinal disorders including ulcerative
 XX colitis, other conditions associated with abnormal vesicle trafficking
 XX including AIDS, allergic reactions, multiple sclerosis and rheumatoid
 XX arthritis. A cell comprising the vector of the invention is useful for
 XX producing non-human transgenic animals. The polynucleotide of the
 XX invention can be used to treat disorders by gene therapy. This sequence
 XX represents one of the isolated NOVX proteins of the invention
 XX
 XX Sequence 446 AA;
 XX
 XX Query Match 89.5%; Score 2116; DB 5; Length 446;
 XX Best Local Similarity 91.0%; Pred. No. 2,2e-192;
 XX Matches 415; Conservative 1; Mismatches 2; Indels 38; Gaps 1;
 XX
 XX 1 MRKISNHSGLRVAVAYPLIGCVGFIYVAYIKMHRATATQAFSTIRAPAGRMGQAH 60
 XX 29 MRKISNHSGLRVAV-----ARMGQAH 50
 XX
 XX 61 SPICGTADGHEHVEYGVIMFDAGSTGTRVHVFQFTRPPREPTLTHERFKAVKPGISAVADD 120
 XX 51 SPLGTADGHEHVEYGVIMFDAGSTGTRVHVFQFTRPPREPTLTHERFKAVKPGISAVADD 110
 XX 121 VSKAOGIRELLDVAKODIPDFWKATPIVVKATAGRLRPGKAQKLLQKVEVFKAAP 180

DB 111 VSKAOGIRELLDVAKODIPDFWKATPIVVKATAGRLRPGKAQKLLQKVEVFKAAP 170
 QY 181 FLVGGDCVSIINNGDEGVSMITINPLTSLKTPGSSVGMILGGSSTOIATLPVEGT 240
 DB 171 FLVGGDCVSIINNGDEGVSMITINPLTSLKTPGSSVGMILGGSSTOIATLPVEGT 230
 QY 241 LQASPPGTLALRMFNRTYKLYSYSLGLGMSARLAILGGVEGQPAKDKELVSPCLSP 300
 DB 231 LQASPPGTLALRMFNRTYKLYSYSLGLGMSARLAILGGVEGQPAKDKELVSPCLSP 290
 QY 301 SFKGEWEHAETTVVSQKAAASLHELCAARVSEVLQNRHREVEVNDVFYAFSTYYDL 360
 DB 291 SFKGEWEHAETTVVSQKAAASLHELCAARVSEVLQNRHREVEVNDVFYAFSTYYDL 350
 QY 361 AAGVGLIDAEKGSLLVYGDPEIAKVCRTLETOPSSPSCMDLTYSLLLOEFGFPRS 420
 DB 351 AAGVGLIDAEKGSLLVYGDPEIAKVCRTLETOPSSPSCMDLTYSLLLOEFGFPRS 410
 QY 421 KVLKLTREKIDNVEFSMALGALFHYIDSLNRQKSPAS 456
 DB 411 KVLKLTREKIDNVEFSMALGALFHYIDSLNRQKSPAS 446
 XX
 XX RESULT 9
 XX AAB72242
 XX ID AAB72242 standard; protein; 471 AA.
 XX AC AAB72242;
 XX DT 14-MAY-2001 (first entry)
 XX DE Mature human CD39 like protein CD39-L2 amino acid sequence.
 XX KW Human CD39-like protein; apyrase; NDPase; platelet function inhibitor;
 XX KW myocardial infarction; cerebral ischemia; angina; arterial thrombosis;
 XX KW cerebral artery thrombosis; platelet aggregation; inflammation;
 XX KW apoptosis; autoimmune disorder; neurological disorder;
 XX KW Alzheimer's disease; Parkinson's disease; cancer; CD39-L2.
 XX OS Homo sapiens.
 XX PN WO200110205-A1.
 XX PD 15-FEB-2001.
 XX PF 09-AUG-2000; 2000WO-US021790.
 XX PR 09-AUG-1999; 99US-00370265.
 XX PR 11-JAN-2000; 2000US-00481238.
 XX PR 25-APR-2000; 2000US-00557800.
 XX PR 26-MAY-2000; 2000US-00583231.
 XX PR 30-JUN-2000; 2000US-00608285.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Ford J, Mulero JJ, Yeung G;
 XX DR N-PSDB; AAF63387.
 XX WPI: 2001-147489/15.
 XX Polynucleotides encoding human CD39-like polypeptides, with apyrase
 XX and/or NDPase activity, which are useful in the treatment of pathological
 XX conditions caused by thrombosis (e.g. myocardial infarction) and
 XX inflammatory disorders.
 XX Claim 53; Page 192-194; 203pp; English.
 XX This invention relates to polynucleotides encoding human CD39-like
 XX polypeptides with apyrase and/or NDPase activity. The polypeptides having
 XX apyrase, including NDPase, activity are useful for inhibiting platelet
 XX function and can therefore be used in the prophylaxis or treatment of
 XX pathological conditions caused by or involving thrombosis or excessive

CC coagulation or excessive platelet aggregation, such as myocardial
 CC infarction, cerebral ischaemia, angina, arterial thrombosis, cerebral
 CC artery thrombosis or intracardiac thrombosis, and conditions associated
 CC with venous thrombosis. CD39-L4 and CD39-L2 polypeptides are useful in
 CC modulating disease states (including platelet aggregation, inflammation
 CC and apoptosis) associated with ADP or other purinergic signalling by
 CC reducing the levels of NTPs. The polypeptides are also useful for
 CC prophylaxis or treatment of inflammation related disorders, such as
 CC disorders involving sepsis or systemic inflammatory response syndrome or
 CC SIRS (and associated conditions such as fever, tachycardia, tachypnea,
 CC cytokine overstimulation); autoimmune disorders such as thrombosis,
 CC atherosclerosis, acute pancreatitis, dermatitis, including psoriasis,
 CC cirrhosis, reperfusion injury, asthma, multiple sclerosis, arthritis;
 CC neurological disorders including neurodegenerative diseases, epilepsy,
 CC depression, Alzheimer's disease, Parkinson's disease, Huntington's
 CC disease, and amyotrophic lateral sclerosis; and cancer. The present
 CC sequence represents mature D39 like protein CD39-L2

XX SQ Sequence 471 AA;

Query Match 84.7%; Score 2003; DB 4; Length 471;
 Best Local Similarity 98.5%; Pred. No. 1.4e-181;
 Matches 388; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MRKISNHGSLRVAKVAYPLGLGVFIYVAYIKWHRATATQAFPSITRAAPGARWQQAH 60
 Db 29 MRKISNHGSLRVAKVAYPLGLGVFIYVAYIKWHRATATQAFPSITRAAPGARWQQAH 88
 Qy 61 SPLGTAADGHEVFYGMFDAGSGTGRVHVFOFTPRPTPTLTHTFFKAVKGLSNAYADD 120
 Db 89 SPLGTAADGHEVFYGMFDAGSGTGRVHVFOFTPRPTPTLTHTFFKAVKGLSNAYADD 148
 Qy 121 VEKSAQGIREDLLVAKQDIPFDFWKATPLVLKATAGLRLLPGEKAQKLQKVEFKASP 180
 Db 149 VEKSAQGIREDLLVAKQDIPFDFWKATPLVLKATAGLRLLPGEKAQKLQKVEFKASP 208
 Qy 181 FLVGDCCVSIIMNGTDEGVSAWITINFLTGLTPGGSSVGMLDLGGGSTQIAFLPRVEGT 240
 Db 209 FLVGDCCVSIIMNGTDEGVSAWITINFLTGLTPGGSSVGMLDLGGGSTQIAFLPRVEGT 268
 Qy 241 LQASPPGYLTALRMFNRVTKLYSYVLGLGLMSARLAILGGVEGQPAKDKGLVSPCLSP 300
 Db 269 LQASPPGYLTALRMFNRVTKLYSYVLGLGLMSARLAILGGVEGQPAKDKGLVSPCLSP 328
 Qy 301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVGVDFVAFYYIDL 360
 Db 329 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVGVDFVAFYYIDL 388
 Qy 361 AAGVGLIDAEKGSLLVVGDPFEIAAKVVCRTLEQ 394
 Db 389 AAGVGLIDAEKGSLLVVGDPFEIAAKVGGSHLER 422

RESULT 10
 AAY44849
 ID AAY44849 standard; protein; 428 AA.
 AC AAY44849;

XX 18-MAY-2000 (first entry)

XX Human CD39-L4 protein.

XX CD39-L4; human; apyrase; nucleotide diphosphatase; NDase;
 KW ATP diphosphohydrolase; ATPase; adenosine diphosphate; ADP; treatment;
 KW platelet aggregation; antithrombotic; thrombosis; myocardial infarction;
 KW cerebral ischaemia; angina; vascular graft; extracorporeal circulation;
 KW molecular weight marker; nutritional supplement; tumour; prevention;
 KW drug targeting; Apyrase Conserved Region; ACR.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1. .22
 FT /label= Leader_peptide
 FT Protein 23. .428
 FT /label= Mature human CD39-L4 protein
 FT /note= "Homologous to human and murine CD39"
 FT Binding-site 54. .58
 FT /label= ATP_Binding_region
 FT Region 129. .134
 FT /label= Apyrase_Conserved_Region
 FT Region 169. .173
 FT /label= Apyrase_Conserved_Region
 FT Region 199. .206
 FT /note= "Conserved motif in ATPDases"
 FN WO200004041-A2.
 PD XX
 PD 27-JAN-2000.
 XX
 PF 16-JUL-1999; 99WO-US016180.
 XX
 PR 16-JUL-1998; 98US-00118205.
 PR 24-JUL-1998; 98US-00122449.
 PR 04-FEB-1999; 99US-00244444.
 PR 19-MAR-1999; 99US-00273447.
 PR 09-JUL-1999; 99US-00350836.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Ford J, Mulero J;
 XX
 XX WPI; 2000-192397/16.
 DR N-PSDB; AAZ50356, AAZ50359.
 XX
 PT New nucleic acid encoding human CD39-like protein, useful for treating
 PT and preventing thrombotic disease.
 XX
 PS Claim 15; Fig 2; 125pp; English.
 XX
 CC The present amino acid sequence is the CD39-L4 protein, an apyrase and/or
 CC nucleotide diphosphatase (NDase). It is isolated from the human foetal
 CC liver-spleen cDNA library, b2HFLS20W. It is a soluble ATP
 CC diphosphohydrolase (ATPDase) and is involved in the hydrolysis of
 CC adenosine diphosphate (ADP), the agonist that causes platelet
 CC aggregation. CD39-L4 protein has 30% and 80% homology to human and murine
 CC CD39. It has platelet aggregation inhibition and antithrombotic activity.
 CC CD39-L4 is used to treat or prevent thrombosis, myocardial infarction,
 CC cerebral ischaemia and angina. It is also used in vitro, to maintain
 CC vascular grafts or during extracorporeal circulation, to hydrolyse NDP,
 CC as molecular weight markers and as nutritional supplements. It is used to
 CC identify therapeutic agents that bind and modulate CD39-L4. It is coupled
 CC to toxins for targeting drugs to tumours or other cells that express CD39
 CC -L4
 XX SQ Sequence 428 AA;
 Query Match 42.3%; Score 999; DB 3; Length 428;
 Best Local Similarity 52.4%; Pred. No. 5.5e-86;
 Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;
 Qy 62 PLGTAADGHEVFYGMFDAGSGTGRVHVFOFT-PPRETPTLTHTFFKAVKGLSNAYADD 120
 Db 40 PINVSA--STLYGIMFDAGSGTGRVHVTFVQKMPGQLPILGEVFDPSVKPGLSAFVDQ 96
 Qy 121 VEKSAQGIREDLLVAKQDIPFDFWKATPLVLKATAGLRLLPGEKAQKLQKVEFKASP 180
 Db 97 PKQAGETVQGLLEVAKQDSIPRSHWKTTPVVLKATAGLRLLPGEKAQKLQKVEFKASP 156
 Qy 181 FLVGDCCVSIIMNGTDEGVSAWITINFLTGLSKTPGGSSVGMLDLGGGSTQIAFLPRVEGT 240
 Db 157 FLVFKGSVSIIMDSGDEGLLAWTVNFLTGLHGRQETVGTLDLGGASTQITFLPQPEKT 216
 Qy 241 LQASPPGYLTALRMFNRVTKLYSYVLGLGLMSARLAILGGVEGQPAKDKGLVSPCLSP 300

Db 217 LEQTPRGVLTSEFMENSTYKLYTHSYLGFGLKAAKRLATLGALETE-GTDGHTFRSACLPR 275
 QY 301 SFKREWEHAETVTVRSQKAAASLHELCAARVSEVLQNRHREDEYVAFPSYYDL 360
 Db 276 WLEAEWIFGVKYQYQGNQSGEVEPEPCYAEVLNRVVGKHLQPEEVORSGSFVAFSYDDR 335
 QY 361 AAGVGLIDAEKGSILVVGDEPEIAKVCRTLETQPOSSPFCMDLTYSLLQD-FGFPFR 419
 Db 336 AVDTMDIDYEKGGILKVEDFERKARVCDNLNENTSSPFLCMDLSYITLALDGFEGFAD 395
 QY 420 SKVTKLTKRIDNVTSMALGALFHYIDSL 448
 Db 396 STVLQTKKVNNIETGMALGATFHLQSL 424
 RESULT 11
 AAB72238
 ID AAB72238 standard; protein; 428 AA.
 XX AAB72238;
 AC
 XX 14-MAY-2001 (first entry)
 DT
 XX Human CD39 like protein CD39-L4 amino acid sequence.
 DE
 XX Human CD39-like protein; apyrase; NDase; platelet function inhibitor;
 KW myocardial infarction; cerebral ischaemia; angina; arterial thrombosis;
 KW cerebral artery thrombosis; platelet aggregation; inflammation;
 KW apoptosis; autoimmune disorder; neurological disorder;
 KW Alzheimer's disease; Parkinson's disease; cancer; CD39-L4.
 KM
 OS Homo sapiens.
 XX
 PN WO200110205-A1.
 XX
 PD 15-FEB-2001.
 XX
 PF 09-AUG-2000; 2000WO-US021790.
 XX
 PR 09-AUG-1999; 99US-00370265.
 PR 11-JAN-2000; 2000US-00481238.
 PR 25-APR-2000; 2000US-00557800.
 PR 26-MAY-2000; 2000US-00583231.
 PR 30-JUN-2000; 2000US-00608285.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Ford J, Mulero JJ, Yeung G;
 XX
 DR WPI; 2001-147489/15.
 DR N-PSDB; AAF63383.
 XX
 PT Polynucleotides encoding human CD39-like polypeptides, with apyrase
 PT and/or NDase activity, which are useful in the treatment of pathological
 PT conditions caused by thrombosis (e.g. myocardial infarction) and
 PT inflammatory disorders.
 XX
 PS Claim 15; Fig 2; 203pp; English.
 XX
 CC This invention relates to polynucleotides encoding human CD39-like
 CC polypeptides with apyrase and/or NDase activity. The polypeptides having
 CC ATPase, including NDase, activity are useful for inhibiting platelet
 CC function and can therefore be used in the prophylaxis or treatment of
 CC pathological conditions caused by or involving thrombosis or excessive
 CC coagulation, cerebral ischaemia, angina, arterial thrombosis, cerebral
 CC infarction, cerebral ischaemia, angina, arterial thrombosis, cerebral
 CC artery thrombosis or intracardiac thrombosis, and conditions associated
 CC with venous thrombosis. CD39-L4 and CD39-L2 polypeptides are useful in
 CC modulating disease states (including platelet aggregation, inflammation
 CC and apoptosis) associated with ADP or other purinergic signalling by
 CC reducing the levels of NDase. The polypeptides are also useful for
 CC prophylaxis or treatment of inflammation related disorders, such as
 CC disorders involving sepsis or systemic inflammatory response syndrome or

CC SIRS (and associated conditions such as fever, tachycardia, tachypnea,
 CC cytokine overstimulation); autoimmune disorders such as thrombosis,
 CC atherosclerosis, acute pancreatitis, dermatitis, including psoriasis,
 CC clithosis, reperfusion injury, asthma, multiple sclerosis, arthritis,
 CC neurological disorders including neurodegenerative diseases, epilepsy,
 CC depression, Alzheimer's disease, Parkinson's disease, Huntington's
 CC disease, and amyotrophic lateral sclerosis; and cancer. The present
 CC sequence represents human CD39 like protein CD39-L4
 CC XX
 SQ Sequence 428 AA;
 Query Match 42.3%; Score 999; DB 4; Length 428;
 Best Local Similarity 52.4%; Pred. No. 5.5e-86;
 Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;
 QY 62 PLGTADGHEVFYQMPDAGSTGTRVHYFOFT-REPREPTLTHFRVAKVPGASVADD 120
 Db 40 PLNVSH---STVYGMFDHSGTGRIRHYTVFQKRPGLPILGEGVPSVAFGLSAFVDQ 96
 QY 121 VEKSAQGIREFLDVAKODIPDFWKAETPLVKATAGLRLPGEKAQKLLQKVEVFKASP 180
 Db 97 PKGAEIVQGLLEVAKDSIPRSHWKTPLVLAATAGLRLPEHKAKALLFEVKEIFRASP 156
 QY 181 FLVGDQCVSINMGTEGVSAMITINFLGSLKTPGSSVGMULDGGSQTQAFPRVSGT 240
 Db 157 FLVPRGVSIVMDGSDGILAVTVNPLTGQLGHRQETVGLDGGASTQTLFPOFEKT 216
 QY 241 LQASPPGYLTALRNRTYKLYSYLYLGLGMSARLALIGVEGQPAKDEKELVSPCLSP 300
 Db 217 LEQTPRGVLTSEFMENSTYKLYTHSYLGFGLKAAKRLATLGALETE-GTDGHTFRSACLPR 275
 QY 301 SFKREWEHAETVTVRSQKAAASLHELCAARVSEVLQNRHREDEYVAFPSYYDL 360
 Db 276 WLEAEWIFGVKYQYQGNQSGEVEPEPCYAEVLNRVVGKHLQPEEVORSGSFVAFSYDDR 335
 QY 361 AAGVGLIDAEKGSILVVGDEPEIAKVCRTLETQPOSSPFCMDLTYSLLQD-FGFPFR 419
 Db 336 AVDTMDIDYEKGGILKVEDFERKARVCDNLNENTSSPFLCMDLSYITLALDGFEGFAD 395
 QY 420 SKVTKLTKRIDNVTSMALGALFHYIDSL 448
 Db 396 STVLQTKKVNNIETGMALGATFHLQSL 424
 RESULT 12
 AAB72243
 ID AAB72243 standard; protein; 428 AA.
 XX AAB72243;
 AC
 XX 14-MAY-2001 (first entry)
 DT
 XX Human CD39 like protein CD39-L4 amino acid sequence.
 DE
 XX Human CD39-like protein; apyrase; NDase; platelet function inhibitor;
 KW myocardial infarction; cerebral ischaemia; angina; arterial thrombosis;
 KW cerebral artery thrombosis; platelet aggregation; inflammation;
 KW apoptosis; autoimmune disorder; neurological disorder;
 KW Alzheimer's disease; Parkinson's disease; cancer; CD39-L4.
 KM
 OS Homo sapiens.
 XX
 PN WO200110205-A1.
 XX
 PD 15-FEB-2001.
 XX
 PF 09-AUG-2000; 2000WO-US021790.
 XX
 PR 09-AUG-1999; 99US-00370265.
 PR 11-JAN-2000; 2000US-00481238.
 PR 25-APR-2000; 2000US-00557800.
 PR 26-MAY-2000; 2000US-00583231.
 PR 30-JUN-2000; 2000US-00608285.
 XX

or chemokine-induced lung injury, inflammatory bowel disease or Crohn's disease. The present sequence is human CD39L4 protein

Sequence 428 AA;

Query Match 42.3%; Score 999; DB 5; Length 428;
Best Local Similarity 52.4%; Pred. No. 5.5e-86;
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

```
QY 62 PLGTADGHEVYVYIMDAGSTGRVAVFOPT-RPRRETPTLTHETFKAVKPGLSAYDD 120
DB 40 PINVSA---STLYGIMFDAGSTGRVAVFOPT-RPRRETPTLTHETFKAVKPGLSAYDD 96
QY 121 VEKSAQIRELIDVAKODIPDFWKATPVLKATAGRLPGEKAKLLQKVEVFKASP 180
DB 97 PKQGAETVQGLLEVAKDSIPRSHWKTTPVLKATAGRLPGEKAKLLQKVEVFKASP 156
QY 181 FLVDDCVSIMNGTDEGVSAWITINFLTGSILKTPGSSVGMLDGGSTOIAFLPRVEGT 240
DB 157 FLVFKGVSIMNGTDEGVSAWITINFLTGSILKTPGSSVGMLDGGSTOIAFLPRVEGT 216
QY 241 LQASPPGYTLALRMNRYTKLYSYGLGMSARLILGVGEQPAKDGKELVSPCLSP 300
DB 217 LEQTPRGVLTSEFNSTYKLYTISYGLGMSARLILGVGEQPAKDGKELVSPCLSP 275
QY 301 SFKGEWHAETVYRVSGQKAAALHDLCAARVSEVLQNRVHTEEVKAVDFYAFSYDDL 360
DB 276 WLEAEWIFGVYKYGQNGEGEVGFPCYAEVLRVKAGKHQPEEVGRGSFYAFSYDDL 335
QY 361 AAGVGLIDAEKGSIVVGDPEIAKVCRTLETQPGSSPCMDLTYSLLQF-RGPR 419
DB 336 AVDTMDIDYEKGILKVEDFERKAEVCDNENFTSGSPFLCMDLSTITALLKDGFGFAD 395
```

RESULT 14
AAY44850
ID AAY44850 standard; protein; 428 AA.

XX AAY44850;

XX AC 18-MAY-2000 (first entry)

XX DE Human CD39-L4 variant-ACR III mutant protein.

XX CD39-L4; human; apyrase; nucleotide diphosphatase; NDPase; variant;
XX ATP diphosphohydrolase; ADPase; adenosine diphosphate; ADP; treatment;
XX platelet aggregation; antithrombotic; thrombosis; myocardial infarction;
XX cerebral ischaemia; angina; vascular graft; extracorporeal circulation;
XX molecular weight marker; nutritional supplement; tumour; prevention;
XX drug targeting; substitution mutation.

XX Homo sapiens.

XX OS Synthetic.

XX Key Location/Qualifiers

XX FT MISC-difference 168 /note= "Wild type Asp substituted with Thr"

XX FT MISC-difference 170 /note= "Wild type Ser substituted with Gln"

XX FT MISC-difference 175 /note= "Wild type Leu substituted with Phe"

XX PN WO200004041-A2.

XX PD 27-JAN-2000.

XX PF 16-JUL-1999; 99WO-US016180.

XX PR 16-JUL-1998; 98US-00118205.

PR 24-JUL-1998; 98US-00122449.
PR 04-FEB-1999; 99US-00244444.
PR 19-MAR-1999; 99US-00273447.
PR 09-JUL-1999; 99US-00350836.

XX (HYSE-) HYSEQ INC.

XX Ford J, Mulero J;

XX MPI: 2000-182397/16.

XX N-PGDB; AA250357.

XX New nucleic acid encoding human CD39-L4 protein, useful for treating and preventing thrombotic disease.

XX Claim 17; Fig 6; 125pp; English.

CC The present amino acid sequence is the CD39-L4 variant, designated as ACR III mutant protein, an apyrase and/or nucleotide diphosphatase (NDPase). It is isolated from the human foetal liver-spleen cDNA library, and is b2HFS20W. It is a soluble ATP diphosphohydrolase (ADPase) and is involved in the hydrolysis of adenosine diphosphate (ADP), the agonist that causes platelet aggregation. CD39-L4 protein has 30% and 80% homology to human and murine CD39. It has platelet aggregation inhibition and antithrombotic activity. CD39-L4 is used to treat or prevent thrombosis, myocardial infarction, cerebral ischaemia and angina. It is also used in vitro, to maintain vascular grafts or during extracorporeal circulation, to hydrolyse NDP, as molecular weight markers and as nutritional supplements. It is used to identify therapeutic agents that bind and modulate CD39-L4. It is coupled to toxins for targeting drugs to tumours or other cells that express CD39-L4

XX Sequence 428 AA;

Query Match 42.1%; Score 996; DB 3; Length 428;
Best Local Similarity 52.4%; Pred. No. 1.1e-85;
Matches 204; Conservative 54; Mismatches 125; Indels 6; Gaps 4;

```
QY 62 PLGTADGHEVYVYIMDAGSTGRVAVFOPT-RPRRETPTLTHETFKAVKPGLSAYDD 120
DB 40 PINVSA---STLYGIMFDAGSTGRVAVFOPT-RPRRETPTLTHETFKAVKPGLSAYDD 96
QY 121 VEKSAQIRELIDVAKODIPDFWKATPVLKATAGRLPGEKAKLLQKVEVFKASP 180
DB 97 PKQGAETVQGLLEVAKDSIPRSHWKTTPVLKATAGRLPGEKAKLLQKVEVFKASP 156
QY 181 FLVDDCVSIMNGTDEGVSAWITINFLTGSILKTPGSSVGMLDGGSTOIAFLPRVEGT 240
DB 157 FLVFKGVSIMNGTDEGVSAWITINFLTGSILKTPGSSVGMLDGGSTOIAFLPRVEGT 216
QY 241 LQASPPGYTLALRMNRYTKLYSYGLGMSARLILGVGEQPAKDGKELVSPCLSP 300
DB 217 LEQTPRGVLTSEFNSTYKLYTISYGLGMSARLILGVGEQPAKDGKELVSPCLSP 275
QY 301 SFKGEWHAETVYRVSGQKAAALHDLCAARVSEVLQNRVHTEEVKAVDFYAFSYDDL 360
DB 276 WLEAEWIFGVYKYGQNGEGEVGFPCYAEVLRVKAGKHQPEEVGRGSFYAFSYDDL 335
QY 361 AAGVGLIDAEKGSIVVGDPEIAKVCRTLETQPGSSPCMDLTYSLLQF-RGPR 419
DB 336 AVDTMDIDYEKGILKVEDFERKAEVCDNENFTSGSPFLCMDLSTITALLKDGFGFAD 395
```

RESULT 15
AAB72240
ID AAB72240 standard; protein; 428 AA.
XX AAB72240;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 09:50:40 ; Search time 49 Seconds
(without alignments)
2618.172 Million cell updates/sec

Title: US-09-905-589A-2

Perfect score: 2364

Sequence: 1 MRKISHNGSLRAKVAAYFLG.....ALGATRYHIDSLNRKSPAS 456

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 28138677 residues

1155919

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :	Published Applications AA:*
1:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2:	/cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3:	/cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6:	/cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
12:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
15:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
16:	/cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17:	/cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2364	100.0	484	13	US-10-092-063-27
2	2364	100.0	484	14	US-10-286-926-27
3	2364	100.0	484	15	US-10-231-913-123
4	2361	99.9	467	15	US-10-231-913-36
5	2357	99.7	484	15	US-10-231-913-124
6	2116	89.5	446	15	US-10-231-913-125
7	2063.5	87.3	455	15	US-10-231-913-125
8	1954	82.7	437	15	US-10-231-913-271
9	1007	42.6	427	15	US-10-231-913-126
10	999	42.3	428	13	US-10-091-085-3
11	999	42.3	428	13	US-10-091-085-3
12	999	42.3	428	13	US-10-092-063-3
13	999	42.3	428	13	US-10-092-063-5
14	999	42.3	428	14	US-10-286-926-3
15	999	42.3	428	14	US-10-286-926-5

16	999	42.3	428	15	US-10-231-913-127	Sequence 127, App
17	996	42.1	428	13	US-10-091-085-7	Sequence 7, Appl1
18	996	42.1	428	13	US-10-092-063-7	Sequence 7, Appl1
19	996	42.1	428	14	US-10-286-926-7	Sequence 2296, Ap
20	992	42.0	428	16	US-10-408-765A-2296	Sequence 25, Appl
21	909	38.5	405	13	US-10-092-063-25	Sequence 25, Appl
22	909	38.5	405	14	US-10-286-926-25	Sequence 25, Appl
23	909	38.5	405	13	US-10-092-063-39	Sequence 876, App
24	823	34.8	330	9	US-09-925-289-876	Sequence 876, App
25	823	34.8	330	10	US-09-925-289-876	Sequence 6447, Ap
26	590.5	25.0	479	15	US-10-369-493-6447	Sequence 19, Appl
27	525.5	22.2	467	9	US-09-129-112-19	Sequence 36762, A
28	508.5	21.5	433	12	US-10-425-114-26762	Sequence 2169, Ap
29	508	21.5	556	9	US-10-369-493-1169	Sequence 2, Appl1
30	506	21.4	462	9	US-09-129-112-2	Sequence 48875, A
31	504	21.3	467	12	US-10-425-114-45875	Sequence 9, Appl1
32	501	21.2	459	9	US-09-129-112-9	Sequence 250158,
33	500	21.2	467	12	US-10-424-599-230158	Sequence 1713, Ap
34	498	21.1	518	15	US-10-369-493-1713	Sequence 45933, A
35	490.5	20.7	443	12	US-10-425-114-45933	Sequence 145076,
36	483	20.4	457	12	US-10-425-114-51762	Sequence 272, App
37	483	20.4	459	12	US-10-424-599-145076	Sequence 15, Appl
38	478.5	20.2	410	15	US-10-231-913-272	Sequence 37241, A
39	463	19.6	462	9	US-10-425-114-37241	Sequence 49932, Ap
40	447	18.9	465	12	US-10-369-493-3333	Sequence 286, App
41	447	18.9	465	15	US-10-425-114-49932	Sequence 74, Appl
42	445	18.8	605	15	US-10-259-165-786	Sequence 283225,
43	443	18.7	455	14	US-10-259-165-786	
44	392	16.6	459	14	US-10-259-165-74	
45	380.5	16.1	540	12	US-10-424-599-283225	

ALIGNMENTS

RESULT 1

US-10-092-063-27

Sequence 27, Application US/10092063

Publication No. US20020173005A1

GENERAL INFORMATION:

APPLICANT: Muleco, Thilo

TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE POLYPEPTIDES

FILE REFERENCE: 28110735908

CURRENT APPLICATION NUMBER: US/10/092, 063

PRIOR FILING DATE: 2002-03-05

PRIOR APPLICATION NUMBER: 09/370, 265

PRIOR FILING DATE: 2002-01-31

PRIOR APPLICATION NUMBER: PCT/US99/16180

PRIOR FILING DATE: 1999-07-16

PRIOR APPLICATION NUMBER: 09/350, 836

PRIOR FILING DATE: 1999-07-09

PRIOR APPLICATION NUMBER: 09/273, 447

PRIOR FILING DATE: 1999-03-19

PRIOR APPLICATION NUMBER: 09/244, 444

PRIOR FILING DATE: 1999-02-04

PRIOR APPLICATION NUMBER: 09/122, 449

PRIOR FILING DATE: 1998-07-24

PRIOR APPLICATION NUMBER: 09/118, 205

PRIOR FILING DATE: 1998-07-16

NUMBER OF SEQ ID NOS: 39

SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 27

LENGTH: 464

TYPE: PRT

ORGANISM: Homo sapiens

US-10-092-063-27

Query Match 100.0%; Score 2364; DB 13; Length 484;

Best Local Similarity 100.0%; Pred. No. 1.4e-231;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 MRKISHNGSLRAKVAAYFLGCVGFIVYAYIKMRHATATQAFSITRAAFARWQQA 60

Db 29 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHEATATQAFSSITRAAPGARWGQAAH 88
QY 61 SPLGTAADGHEVYIGIMFDAGSGTGRVHVFOFTRPPRETPTLTHETFKAVKGLSAYADD 120
Db 89 SPLGTAADGHEVYIGIMFDAGSGTGRVHVFOFTRPPRETPTLTHETFKAVKGLSAYADD 148
QY 121 VEKSAQIGIRELLDVAKQDIPDFWKATPLVLKATAGRLLLPGEKAQKLLQKVEVFKASP 180
Db 149 VEKSAQIGIRELLDVAKQDIPDFWKATPLVLKATAGRLLLPGEKAQKLLQKVEVFKASP 208
QY 181 FLVGDDCVSINMGTDGVSAMITINFLTSLKTPGGSSVGMDDLGGSGTQIAFLPRVEGT 240
Db 209 FLVGDDCVSINMGTDGVSAMITINFLTSLKTPGGSSVGMDDLGGSGTQIAFLPRVEGT 268
QY 241 LQASPPGYLTALRMFNRTYKLYSYSLGGLMSARLAILGGVEGQPAKDGKELVSPCLSP 300
Db 269 LQASPPGYLTALRMFNRTYKLYSYSLGGLMSARLAILGGVEGQPAKDGKELVSPCLSP 328
QY 301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTTEEVKHVDVFAFSYYIDL 360
Db 329 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTTEEVKHVDVFAFSYYIDL 388
QY 361 AAGVGLIDAEGKGLVVGDFEIAAKYVCRILETQPOSSPFCMDLTYVSLLLQEFGPFRS 420
Db 389 AAGVGLIDAEGKGLVVGDFEIAAKYVCRILETQPOSSPFCMDLTYVSLLLQEFGPFRS 448
QY 421 KVLKLTBKIDNVETSWALGAIFHYIDSINRQKSPAS 456
Db 449 KVLKLTBKIDNVETSWALGAIFHYIDSINRQKSPAS 484

RESULT 2

US-10-286-926-27
; Sequence 27, Application US/10286926
; Publication No. US2003017552A1
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; APPLICANT: Yeung, George
; TITLE OF INVENTION: Methods and Materials Relating to CD39-Like
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 28110/36457CON
; CURRENT APPLICATION NUMBER: US/10/286,926
; CURRENT FILING DATE: 2002-11-01
; PRIOR APPLICATION NUMBER: 09/557,800
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/481,238
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 09/370,265
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/122449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/244444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 27
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-286-926-27

Query Match 100.0%; Score 2364; DB 14; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.4e-231;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHEATATQAFSSITRAAPGARWGQAAH 60
Db 29 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHEATATQAFSSITRAAPGARWGQAAH 88
QY 61 SPLGTAADGHEVYIGIMFDAGSGTGRVHVFOFTRPPRETPTLTHETFKAVKGLSAYADD 120
Db 89 SPLGTAADGHEVYIGIMFDAGSGTGRVHVFOFTRPPRETPTLTHETFKAVKGLSAYADD 148
QY 121 VEKSAQIGIRELLDVAKQDIPDFWKATPLVLKATAGRLLLPGEKAQKLLQKVEVFKASP 180
Db 149 VEKSAQIGIRELLDVAKQDIPDFWKATPLVLKATAGRLLLPGEKAQKLLQKVEVFKASP 208
QY 181 FLVGDDCVSINMGTDGVSAMITINFLTSLKTPGGSSVGMDDLGGSGTQIAFLPRVEGT 240
Db 209 FLVGDDCVSINMGTDGVSAMITINFLTSLKTPGGSSVGMDDLGGSGTQIAFLPRVEGT 268
QY 241 LQASPPGYLTALRMFNRTYKLYSYSLGGLMSARLAILGGVEGQPAKDGKELVSPCLSP 300
Db 269 LQASPPGYLTALRMFNRTYKLYSYSLGGLMSARLAILGGVEGQPAKDGKELVSPCLSP 328
QY 301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTTEEVKHVDVFAFSYYIDL 360
Db 329 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTTEEVKHVDVFAFSYYIDL 388
QY 361 AAGVGLIDAEGKGLVVGDFEIAAKYVCRILETQPOSSPFCMDLTYVSLLLQEFGPFRS 420
Db 389 AAGVGLIDAEGKGLVVGDFEIAAKYVCRILETQPOSSPFCMDLTYVSLLLQEFGPFRS 448
QY 421 KVLKLTBKIDNVETSWALGAIFHYIDSINRQKSPAS 456
Db 449 KVLKLTBKIDNVETSWALGAIFHYIDSINRQKSPAS 484

RESULT 3

US-10-231-913-123
; Sequence 123, Application US/10231913
; Publication No. US20040005576A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia S.
; APPLICANT: Li, Li
; APPLICANT: Fatturajan, Meera
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Sureeh G.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Edinger, Schlomit
; APPLICANT: Peyman, John A.
; APPLICANT: Stone, David J.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Boldog, Ference L.
; APPLICANT: Colman, Steven D.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Zethusen, Bryan D.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-216
; CURRENT APPLICATION NUMBER: US/10/231,913
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 60/251,660
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: 60/255,029
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/260,326
; PRIOR FILING DATE: 2001-01-08

PRIOR APPLICATION NUMBER: 60/263,800
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/269,942
PRIOR FILING DATE: 2001-02-20
PRIOR APPLICATION NUMBER: 60/286,183
PRIOR FILING DATE: 2001-04-24
PRIOR APPLICATION NUMBER: 60/313,627
PRIOR FILING DATE: 2001-08-20
PRIOR APPLICATION NUMBER: 60/318,712
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 123
LENGTH: 484
TYPE: PRT
ORGANISM: Homo sapiens
US-10-231-913-123

Query Match 100.0%; Score 2364; DB 15; Length 484;
Best Local Similarity 100.0%; Pred. No. 1,4e+231;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRKISNHSRLRAKAYAPLGLCVGFYVAYIKMHRATATQAFSITRAAPGARWGQAH 60
29 MRKISNHSRLRAKAYAPLGLCVGFYVAYIKMHRATATQAFSITRAAPGARWGQAH 88
61 SPLGTADGHEVFGYIMPDAGSTGRVHVFQTRPREPTLTHETFKAVKPGLSAYAD 120
89 SPLGTADGHEVFGYIMPDAGSTGRVHVFQTRPREPTLTHETFKAVKPGLSAYAD 148
121 VEKSAQGIKRELLDVAKODIPDFWKATPLVKATAGRLPGEKQKLLQKXEVFKASP 180
149 VEKSAQGIKRELLDVAKODIPDFWKATPLVKATAGRLPGEKQKLLQKXEVFKASP 208
181 FLVGGDDCVSINMGTDGVSAMITINFLGSLKTPGSSVGMGLDGGSTQIAFLPRVEGT 240
209 FLVGGDDCVSINMGTDGVSAMITINFLGSLKTPGSSVGMGLDGGSTQIAFLPRVEGT 268
241 LQASPPGYLTALRMFRYTKLYSYSLGLGMSARLALIGVEQOPKDGKELVSPCLSP 300
269 LQASPPGYLTALRMFRYTKLYSYSLGLGMSARLALIGVEQOPKDGKELVSPCLSP 328
301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHTEEVKVDYFAFSYYDL 360
329 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHTEEVKVDYFAFSYYDL 388
361 AAGVGLIDAEKGSILVGDPEIATKVCRTLETQPOSSPFCMDLTYVSLLLQEFSPRS 420
389 AAGVGLIDAEKGSILVGDPEIATKVCRTLETQPOSSPFCMDLTYVSLLLQEFSPRS 448
421 KYLKLTRKIDNVETSWALGAIFFHYIDSINRQKSPAS 456
449 KYLKLTRKIDNVETSWALGAIFFHYIDSINRQKSPAS 484

RESULT 4
US-10-231-913-36
Sequence 36, Application US/10231913
Publication No. US20040005576A1
GENERAL INFORMATION:
APPLICANT: Guo, Xiaojia S.
APPLICANT: Li, Li
APPLICANT: Patuturajan, Meera
APPLICANT: Shimkets, Richard A.
APPLICANT: Casman, Stacie J.
APPLICANT: Malvankar, Uriel M.
APPLICANT: Tchernev, Velizar T.
APPLICANT: Verneet, Corine A.
APPLICANT: Spytek, Kimberly A.
APPLICANT: Shenoy, Suresh G.
APPLICANT: Alsebrook II, John P.
APPLICANT: Edinger, Schlomit
APPLICANT: Peyman, John A.

APPLICANT: Stone, David J.
APPLICANT: Ellerman, Karen
APPLICANT: Gangoli, Esna A.
APPLICANT: Boldog, Ference L.
APPLICANT: Coleman, Steven D.
APPLICANT: Eileen, Andrew J.
APPLICANT: Liu, Xiaohong
APPLICANT: Padigaru, Muralidhara
APPLICANT: Spaderna, Steven K.
APPLICANT: Zetrisen, Bryan D.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-216
CURRENT APPLICATION NUMBER: US/10/231,913
CURRENT FILING DATE: 2002-08-30
PRIOR APPLICATION NUMBER: 60/251,660
PRIOR FILING DATE: 2000-12-06
PRIOR APPLICATION NUMBER: 60/255,029
PRIOR FILING DATE: 2000-12-12
PRIOR APPLICATION NUMBER: 60/260,326
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 60/263,800
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/269,942
PRIOR FILING DATE: 2001-02-20
PRIOR APPLICATION NUMBER: 60/286,183
PRIOR FILING DATE: 2001-04-24
PRIOR APPLICATION NUMBER: 60/313,627
PRIOR FILING DATE: 2001-08-20
PRIOR APPLICATION NUMBER: 60/318,712
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 36
LENGTH: 467
TYPE: PRT
ORGANISM: Homo sapiens
US-10-231-913-36

Query Match 99.9%; Score 2361; DB 15; Length 467;
Best Local Similarity 99.8%; Pred. No. 2,6e+231;
Matches 455; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 MRKISNHSRLRAKAYAPLGLCVGFYVAYIKMHRATATQAFSITRAAPGARWGQAH 60
12 MRKISNHSRLRAKAYAPLGLCVGFYVAYIKMHRATATQAFSITRAAPGARWGQAH 71
61 SPLGTADGHEVFGYIMPDAGSTGRVHVFQTRPREPTLTHETFKAVKPGLSAYAD 120
72 SPLGTADGHEVFGYIMPDAGSTGRVHVFQTRPREPTLTHETFKAVKPGLSAYAD 131
121 VEKSAQGIKRELLDVAKODIPDFWKATPLVKATAGRLPGEKQKLLQKXEVFKASP 180
132 VEKSAQGIKRELLDVAKODIPDFWKATPLVKATAGRLPGEKQKLLQKXEVFKASP 191
181 FLVGGDDCVSINMGTDGVSAMITINFLGSLKTPGSSVGMGLDGGSTQIAFLPRVEGT 240
192 FLVGGDDCVSINMGTDGVSAMITINFLGSLKTPGSSVGMGLDGGSTQIAFLPRVEGT 251
241 LQASPPGYLTALRMFRYTKLYSYSLGLGMSARLALIGVEQOPKDGKELVSPCLSP 300
252 LQASPPGYLTALRMFRYTKLYSYSLGLGMSARLALIGVEQOPKDGKELVSPCLSP 311
301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHTEEVKVDYFAFSYYDL 360
312 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHTEEVKVDYFAFSYYDL 371
361 AAGVGLIDAEKGSILVGDPEIATKVCRTLETQPOSSPFCMDLTYVSLLLQEFSPRS 420
372 AAGVGLIDAEKGSILVGDPEIATKVCRTLETQPOSSPFCMDLTYVSLLLQEFSPRS 431
421 KYLKLTRKIDNVETSWALGAIFFHYIDSINRQKSPAS 456
432 KYLKLTRKIDNVETSWALGAIFFHYIDSINRQKSPAS 467

```
RESULT 5
US-10-231-913-124
; Sequence 124, Application US/10231913
; Publication No. US20040005576A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia S.
; APPLICANT: Li, Li
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Edinger, Schlomit
; APPLICANT: Peyman, John A.
; APPLICANT: Stone, David J.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Boldog, Ference L.
; APPLICANT: Colman, Steven D.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-216
; CURRENT APPLICATION NUMBER: US/10/231,913
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 60/251,660
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: 60/255,029
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/260,326
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/263,800
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/269,942
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/318,712
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 124
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-231-913-124

Query Match          99.78; Score 2357; DB 15; Length 484;
Best Local Similarity 99.68; Pred. No. 7e-231;
Matches 454; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKTISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFSTTRAAPGARWQQAH 60
Db 29 MKTISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFSTTRAAPGARWQQAH 88

Qy 61 SPLGTADAGHEVFYVINGMFDAGSTGTRVHVQFTRPPREPTLTTHETFKAVKGLSAYADD 120
Db 89 SPLGTADAGHEVFYVINGMFDAGSTGTRVHVQFTRPPREPTLTTHETFKAVKGLSAYADD 148

Qy 121 VEKSAQGIREDLLDVAQKODIFFDFWPKATPLVLKATAGLRLLPGKAKLQKVKVEFKASP 180
Db 149 VEKSAQGIREDLLDVAQKODIFFDFWPKATPLVLKATAGLRLLPGKAKLQKVKVEFKASP 208
```

```
181 FLVGDDCVSIMNGTDEGVSAMITINFLTGSLKTPGSSVGMGLDLGGGSTQIAFLPRVEGT 240
209 FLVGDDCVSIMNGTDEGVSAMITINFLTGSLKTPGSSVGMGLDLGGGSTQIAFLPRVEGT 268
241 LQASPPGYLTALRMFNRTYKLYSYVGLGIMSGARLAILGGVEGQPAKDGKELVSPCLSP 300
269 LQASPPGYLTALRMFNRTYKLYSYVGLGIMSGARLAILGGVEGQPAKDGKELVSPCLSP 328
301 SPKGWEHAEVTVRVSGQKAAASHLHCAARVSEVLQNRVHRTTEEVGHVDFYAFSYYDL 360
329 SPKGWEHAEVTVRVSGQKAAASHLHCAARVSEVLQNRVHRTTEEVGHVDFYAFSYYDL 388
361 AAGVGLIDARKGSLVVGDFEIIAAKYVCRLETQPOSSPSCMDLTYVSLLLQEFEGPERS 420
389 AAGVGLIDARKGSLVVGDFEIIAAKYVCRLETQPOSSPSCMDLTYVSLLLQEFEGPERS 448
421 KVLKLTTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456
449 KVLKLTTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 484

RESULT 6
US-10-231-913-38
; Sequence 38, Application US/10231913
; Publication No. US20040005576A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia S.
; APPLICANT: Li, Li
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Edinger, Schlomit
; APPLICANT: Peyman, John A.
; APPLICANT: Stone, David J.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Boldog, Ference L.
; APPLICANT: Colman, Steven D.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-216
; CURRENT APPLICATION NUMBER: US/10/231,913
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 60/251,660
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: 60/255,029
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/260,326
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/263,800
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/269,942
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/286,183
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/313,627
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/318,712
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
```

LENGTH: 446
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-231-913-38

Query Match 89.5%; Score 2116; DB 15; Length 446;
 Best Local Similarity 91.0%; Pred. No. 2,2e-206;
 Matches 415; Conservative 1; Mismatches 2; Indels 38; Gaps 1;

```

QY 1 MKKISNHSGLRVKAVYPLGLCVGVFIYVAYIKMHRATATQAFPSITRAPAGAWGQAH 60
DB 29 MKKISNHSGLRVAV-----ARWQAH 50
QY 61 SPLGTADGHEVYFGIMFDAGSTGTGVHVVFQTRPRETPTLTHTETPKAVKGLSAYAD 120
DB 51 SPLGTADGHEVYFGIMFDAGSTGTGVHVVFQTRPRETPTLTHTETPKAVKGLSAYAD 110
QY 121 VEKSAQGIREFLLVAQODIFPDFWKATPLVLTAKATAGRLPGEKAQKLQKVEVFKA 180
DB 111 VEKSAQGIREFLLVAQODIFPDFWKATPLVLTAKATAGRLPGEKAQKLQKVEVFKA 170
QY 181 FLVGDCCVSIINMGDEGVSAWITINFLTGSLSKTPGSSVGMULDGGGSTQIAFLPREV 240
DB 171 FLVGDCCVSIINMGDEGVSAWITINFLTGSLSKTPGSSVGMULDGGGSTQIAFLPREV 230
QY 241 LQASPPGYLTALRMNRTYLYSYSLGLGMSARLAILGVEGQPAKDGKELVSPCLSP 300
DB 231 LQASPPGYLTALRMNRTYLYSYSLGLGMSARLAILGVEGQPAKDGKELVSPCLSP 290
QY 301 SFKGEWEHAETVYRVSQKAAASLHELCAARVSEVLQNRVHRTVEYKVDYFAFSYYDL 360
DB 291 SFKGEWEHAETVYRVSQKAAASLHELCAARVSEVLQNRVHRTVEYKVDYFAFSYYDL 350
QY 361 AAGVGLIDAEKGGSLVVDDEFIATKYVCRTEFTQPOSSPCMDLTYVSLILOEFGFP 420
DB 351 AAGVGLIDAEKGGSLVVDDEFIATKYVCRTEFTQPOSSPCMDLTYVSLILOEFGFP 410
QY 421 KVLKLTIRKIDIVETSMALGAIFFHYIDSINROKSPAS 456
DB 411 KVLKLTIRKIDIVETSMALGAIFFHYIDSINROKSPAS 446

```

RESULT 7

US-10-231-913-125
 Sequence 125, Application US/10231913
 Publication No. US20040005576A1

GENERAL INFORMATION:

APPLICANT: Guo, Xiaojia S.

APPLICANT: Li, Li

APPLICANT: Patuturajan, Meera

APPLICANT: Shinkets, Richard A.

APPLICANT: Caeman, Stacie J.

APPLICANT: Malvankar, Uriel M.

APPLICANT: Tchenev, Velizar I.

APPLICANT: Verneel, Corine A.

APPLICANT: Spytek, Kimberly A.

APPLICANT: Shenoy, Suresh G.

APPLICANT: Alebrock II, John P.

APPLICANT: Edinger, Schlomit

APPLICANT: Peyman, John A.

APPLICANT: Stone, David J.

APPLICANT: Ellerman, Karen

APPLICANT: Gangoli, Esha A.

APPLICANT: Boldog, Ferenc L.

APPLICANT: Colman, Steven D.

APPLICANT: Eisen, Andrew J.

APPLICANT: Liu, Xiaohong

APPLICANT: Padigaru, Muralidhara

APPLICANT: Spaderna, Steven K.

APPLICANT: Zernusen, Bryan D.

APPLICANT: Caeman, Stacie J.

APPLICANT: Malvankar, Uriel M.

APPLICANT: Caeman, Stacie J.

APPLICANT: Malvankar, Uriel M.

CURRENT FILING DATE: 2002-08-30
 PRIOR APPLICATION NUMBER: 60/251,660
 PRIOR FILING DATE: 2000-12-06
 PRIOR APPLICATION NUMBER: 60/255,029
 PRIOR FILING DATE: 2000-12-12
 PRIOR APPLICATION NUMBER: 60/260,326
 PRIOR FILING DATE: 2001-01-08
 PRIOR APPLICATION NUMBER: 60/263,800
 PRIOR FILING DATE: 2001-01-24
 PRIOR APPLICATION NUMBER: 60/269,942
 PRIOR FILING DATE: 2001-02-20
 PRIOR APPLICATION NUMBER: 60/266,183
 PRIOR FILING DATE: 2001-04-24
 PRIOR APPLICATION NUMBER: 60/313,627
 PRIOR FILING DATE: 2001-08-20
 PRIOR APPLICATION NUMBER: 60/318,712
 PRIOR FILING DATE: 2001-09-12
 NUMBER OF SEQ ID NOS: 292
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 125
 LENGTH: 455
 TYPE: PRT
 ORGANISM: Rattus norvegicus
 US-10-231-913-125

Query Match 87.3%; Score 2063.5; DB 15; Length 455;
 Best Local Similarity 86.6%; Pred. No. 5.2e-201;
 Matches 394; Conservative 27; Mismatches 33; Indels 1; Gaps 1;

```

QY 1 MKKISNHSGLRVKAVYPLGLCVGVFIYVAYIKMHRATATQAFPSITRAPAGAWGQAH 60
DB 1 MKKISNHSGLRVKAVYPLGLCVGVFIYVAYIKMHRATATQAFPSITRAPAGAWGQAH 60
QY 61 SPLGTADGHEVYFGIMFDAGSTGTGVHVVFQTRPRETPTLTHTETPKAVKGLSAYAD 120
DB 61 SPLGTADGHEVYFGIMFDAGSTGTGVHVVFQTRPRETPTLTHTETPKAVKGLSAYAD 120
QY 121 VEKSAQGIREFLLVAQODIFPDFWKATPLVLTAKATAGRLPGEKAQKLQKVEVFKA 180
DB 121 VEKSAQGIREFLLVAQODIFPDFWKATPLVLTAKATAGRLPGEKAQKLQKVEVFKA 180
QY 181 FLVGDCCVSIINMGDEGVSAWITINFLTGSLSKTPGSSVGMULDGGGSTQIAFLPREV 240
DB 181 FLVGDCCVSIINMGDEGVSAWITINFLTGSLSKTPGSSVGMULDGGGSTQIAFLPREV 240
QY 241 LQASPPGYLTALRMNRTYLYSYSLGLGMSARLAILGVEGQPAKDGKELVSPCLSP 300
DB 241 LQASPPGYLTALRMNRTYLYSYSLGLGMSARLAILGVEGQPAKDGKELVSPCLSP 300
QY 301 SFKGEWEHAETVYRVSQKAAASLHELCAARVSEVLQNRVHRTVEYKVDYFAFSYYDL 360
DB 301 SFKGEWEHAETVYRVSQKAAASLHELCAARVSEVLQNRVHRTVEYKVDYFAFSYYDL 359
QY 361 AAGVGLIDAEKGGSLVVDDEFIATKYVCRTEFTQPOSSPCMDLTYVSLILOEFGFP 420
DB 360 AAGVGLIDAEKGGSLVVDDEFIATKYVCRTEFTQPOSSPCMDLTYVSLILOEFGFP 419
QY 421 KVLKLTIRKIDIVETSMALGAIFFHYIDSINROKSPAS 455
DB 420 KVLKLTIRKIDIVETSMALGAIFFHYIDSINROKSPAS 454

```

RESULT 8

US-10-231-913-271
 Sequence 271, Application US/10231913
 Publication No. US20040005576A1

GENERAL INFORMATION:

APPLICANT: Guo, Xiaojia S.

APPLICANT: Li, Li

APPLICANT: Patuturajan, Meera

APPLICANT: Shinkets, Richard A.

APPLICANT: Caeman, Stacie J.

APPLICANT: Malvankar, Uriel M.

APPLICANT: Tchernev, Velizar T.
APPLICANT: Vernet, Corine A.
APPLICANT: Spytek, Kimberly A.
APPLICANT: Shenoy, Suresh G.
APPLICANT: Alsobrook II, John P.
APPLICANT: Edinger, Schlomit
APPLICANT: Peyman, John A.
APPLICANT: Stone, David J.
APPLICANT: Ellerman, Karen
APPLICANT: Gangolli, Esha A.
APPLICANT: Boldog, Ference L.
APPLICANT: Colman, Steven D.
APPLICANT: Eisen, Andrew J.
APPLICANT: Liu, Xiaohong
APPLICANT: Padigaru, Muralidhara
APPLICANT: Spaderna, Steven K.
APPLICANT: Zerhusen, Bryan D.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-216
CURRENT APPLICATION NUMBER: US/10/231,913
CURRENT FILING DATE: 2002-08-30
PRIOR APPLICATION NUMBER: 60/251,660
PRIOR FILING DATE: 2000-12-06
PRIOR APPLICATION NUMBER: 60/255,029
PRIOR FILING DATE: 2000-12-12
PRIOR APPLICATION NUMBER: 60/260,326
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 60/263,800
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/269,942
PRIOR FILING DATE: 2001-02-20
PRIOR APPLICATION NUMBER: 60/286,183
PRIOR FILING DATE: 2001-04-24
PRIOR APPLICATION NUMBER: 60/313,627
PRIOR FILING DATE: 2001-08-20
PRIOR APPLICATION NUMBER: 60/318,712
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 271
LENGTH: 379
TYPE: PRT
ORGANISM: Homo sapiens
US-10-231-913-271

Query Match 82.7%; Score 1954; DB 15; Length 379;
Best Local Similarity 99.7%; Pred. No. 5.6e-190;
Matches 378; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 71 EVFYGMFDAGSTGRVHVQFTRPPREPTTLTHETFKAVKPGLSAYADDVEKSAQIGRE 130
Db 1 EVFYGMFDAGSTGRVHVQFTRPPREPTTLTHETFKAVKPGLSAYADDVEKSAQIGRE 60
Qy 131 LLDVAKQDIPFDWKATPLVKATAGLRLLPGKAQKLLQKVEFKASPFVLGDDCVSI 190
Db 61 LLDVAKQDIPFDWKATPLVKATAGLRLLPGKAQKLLQKVEFKASPFVLGDDCVSI 120
Qy 191 MNGTDEGVSAMTINFLTGSLKTPGSSVGMPLDGGSTQIAFLPVEGTLOASPPGYLT 250
Db 121 MNGTDEGVSAMTINFLTGSLKTPGSSVGMPLDGGSTQIAFLPVEGTLOASPPGYLT 180
Qy 251 ALRMFNRTYKLSVSYGLGLMGARLAILGGVGGQPKGKELVSPCLSPSPKGEWEHAE 310
Db 181 ALRMFNRTYKLSVSYGLGLMGARLAILGGVGGQPKGKELVSPCLSPSPKGEWEHAE 240
Qy 311 VTYRVSQKAAASLHELCAARVSEVLQNRVHRTTEVKVDFYAFSYYYDLAAGVGLIDAE 370
Db 241 VTYRVSQKAAASLHELCAARVSEVLQNRVHRTTEVKVDFYAFSYYYDLAAGVGLIDAE 300
Qy 371 KGSLLVGDPEIAAKYVCRTLETQPOSSPSCMDLTIVSLLIQEFPFPPSKVLKTRKID 430
Db 301 KGSLLVGDPEIAAKYVCRTLETQPOSSPSCMDLTIVSLLIQEFPFPPSKVLKTRKID 360

Qy 431 NVETSWALGAIFHYIDSIN 449
Db 361 NVETSWALGAIFHYIDSIN 379

RESULT 9
US-10-231-913-126
; Sequence 126, Application US/10231913
; Publication No. US20040005576A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia S.
; APPLICANT: Li, Li
; APPLICANT: Patturajan, Meera
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Edinger, Schlomit
; APPLICANT: Peyman, John A.
; APPLICANT: Stone, David J.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Boldog, Ference L.
; APPLICANT: Colman, Steven D.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Zerhusen, Bryan D.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-216
CURRENT APPLICATION NUMBER: US/10/231,913
CURRENT FILING DATE: 2002-08-30
PRIOR APPLICATION NUMBER: 60/251,660
PRIOR FILING DATE: 2000-12-06
PRIOR APPLICATION NUMBER: 60/255,029
PRIOR FILING DATE: 2000-12-12
PRIOR APPLICATION NUMBER: 60/260,326
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 60/263,800
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/269,942
PRIOR FILING DATE: 2001-02-20
PRIOR APPLICATION NUMBER: 60/286,183
PRIOR FILING DATE: 2001-04-24
PRIOR APPLICATION NUMBER: 60/313,627
PRIOR FILING DATE: 2001-08-20
PRIOR APPLICATION NUMBER: 60/318,712
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 126
LENGTH: 427
TYPE: PRT
ORGANISM: Mus musculus
US-10-231-913-126

Query Match 42.6%; Score 1007; DB 15; Length 427;
Best Local Similarity 50.0%; Pred. No. 2.2e-93;
Matches 213; Conservative 62; Mismatches 133; Indels 18; Gaps 6;

Qy 37 ATATQAFSTTRAAPGA-----RWGQ-----QAHSPLGTADGHEVFFGIMFDAGSTG 84
Db 2 ATSGAVFMILLIACVSTVFYREQQTFEGVFLSSMCPINVSAG---TFYGINFDAGSTG 58
Qy 85 TRVHVFOFT-RPPREPTTLTHETFKAVKPGLSAYADDVEKSAQIGRELLDVAKQDIPFD 143
Db 59 TRIHVITFVQKTAGQLPFLGEIFDSVKPGLSAFVDPQKQGAETVQELLEVAKDSIPRSH 118

QY 144 WKATPLVTKATAGTLLPGEKAKLLOKKEVFKASPELVGDCCVIMNCTDEGSAMIT 203
 DB 119 WERPEVVKATAGTLLPGEKAKLLOKKEVFKASPELVGDCCVIMNCTDEGSAMIT 178
 QY 204 INFLTGSILKTPGSSVGMGLDGGSTQIAFLPRVEGTLOASPPGYLTALMERNRYKYS 263
 DB 179 VNFPLTGLHORGQETVGTLLDGGASTQITFLPGEKTELEOTPPGYLTALMERNRYKYS 238
 QY 264 YSYGLGLMSARLALIGCVSGQPKADKELVSPCLSPFGEMENHAEVYTRVSGQAAAS 323
 DB 239 HSYLGFGLKARLALIGCVSGQPKADKELVSPCLSPFGEMENHAEVYTRVSGQAAAS 297
 QY 324 LHEICARVEVLONRVHRTVEVGVDFVAFSYYYDLAAGVGLDAEKGSILVGDPELA 383
 DB 298 GFPEVYAEVLRVYQKHLHQBEPVRSFAFYFYDDAADTHLIDYKGVLVKVEDFERK 357
 QY 384 AKYVCRLETOPQSSPSCMDLYVSLLOE-FGPPSKYLKTRKIDNVEISMALGATF 442
 DB 358 AREVCDMLGSSSGSPFLCMDLYTALTKDGFADGTLLOLKVNMIETGMALGATF 417
 QY 443 HYIDSL 448
 DB 418 HLQSL 423

RESULT 10

US-10-091-085-3
 / Sequence 3, Application US/10091085
 / Publication No. US20020146772A1
 / GENERAL INFORMATION:
 / APPLICANT: Mulero, Julio
 / TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
 / TITLE OF INVENTION: POLYPEPTIDES
 / FILE REFERENCE: 28110/35761
 / CURRENT APPLICATION NUMBER: US/10/091,085
 / PRIOR FILING DATE: 2002-03-05
 / PRIOR APPLICATION NUMBER: 09/350,836
 / PRIOR FILING DATE: 1999-07-09
 / PRIOR APPLICATION NUMBER: 09/273,447
 / PRIOR FILING DATE: 1999-03-19
 / PRIOR APPLICATION NUMBER: 09/118,205
 / PRIOR FILING DATE: 1998-07-16
 / PRIOR APPLICATION NUMBER: 09/122,449
 / PRIOR FILING DATE: 1998-07-24
 / PRIOR APPLICATION NUMBER: 09/244,444
 / PRIOR FILING DATE: 1999-02-04
 / NUMBER OF SEQ ID NOS: 23
 / SOFTWARE: PatentIn Ver. 2.0
 / SEQ ID NO 3
 / LENGTH: 428
 / TYPE: PRT
 / ORGANISM: Homo sapiens
 / US-10-091-085-3

Query Match 42.3%; Score 999; DB 13; Length 428;
 Best Local Similarity 52.4%; Pred. No. 1.4e-92;
 Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;
 QY 62 PLGTADGHEVYVYGMFDAGSTGRVHVFQPT-RPPREPTLTHETFAVKPGLSAYADD 120
 DB 40 PINVSA---STLYGMFDAGSTGRVHVFQPT-RPPREPTLTHETFAVKPGLSAYADD 96
 QY 121 VEKSAQIRELLDVAKODIPDFWKATPLVTKATAGTLLPGEKAKLLOKKEVFKASP 180
 DB 97 PRGAEVTVQGLLEVAKODIPDFWKATPLVTKATAGTLLPGEKAKLLOKKEVFKASP 156
 QY 181 FLVGDCCVIMNCTDEGSAMITINFLTGSILKTPGSSVGMGLDGGSTQIAFLPRVEGT 240
 DB 157 FLVPGKGSVIMKDSDEGLAWTVNVLGQLHGRQETVGTLLDGGASTQITFLPGEKTE 216
 QY 241 LQASPGVLTALMERNRYKYSYVGLGLMSARLALIGCVSGQPKADKELVSPCLSP 300

DB 217 LEQTPRGVLTSEFEMENSTYKLYTHSYLGFGLKARLALIGCVSGQPKADKELVSPCLSP 275
 QY 301 SFKGEWENHAEVYTRVSGQAAASLHEICARVEVLONRVHRTVEVGVDFVAFSYYYDL 360
 DB 276 WLEAEWIVGVKYYQGGNQBEGVGFPEVYAEVLRVYQKHLHQBEPVRSFAFYFYDDAADTHL 335
 QY 361 AAGVGLDAEKGSILVGDPELAAYVCRLETOPQSSPSCMDLYVSLLOE-FGPPR 419
 DB 336 AVDTMIDYKGVLTSEFEMENSTYKLYTHSYLGFGLKARLALIGCVSGQPKADKELVSPCLSP 395
 QY 420 SKYTLKTRKIDNVEISMALGATFHYIDSL 448
 DB 396 STVLOLTKVNMIETGMALGATFHLQSL 424

RESULT 11

US-10-091-085-5
 / Sequence 5, Application US/10091085
 / Publication No. US20020146772A1
 / GENERAL INFORMATION:
 / APPLICANT: Mulero, Julio
 / TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
 / TITLE OF INVENTION: POLYPEPTIDES
 / FILE REFERENCE: 28110/35761
 / CURRENT APPLICATION NUMBER: US/10/091,085
 / PRIOR FILING DATE: 2002-03-05
 / PRIOR APPLICATION NUMBER: 09/350,836
 / PRIOR FILING DATE: 1999-07-09
 / PRIOR APPLICATION NUMBER: 09/273,447
 / PRIOR FILING DATE: 1999-03-19
 / PRIOR APPLICATION NUMBER: 09/118,205
 / PRIOR FILING DATE: 1998-07-16
 / PRIOR APPLICATION NUMBER: 09/122,449
 / PRIOR FILING DATE: 1998-07-24
 / PRIOR APPLICATION NUMBER: 09/244,444
 / PRIOR FILING DATE: 1999-02-04
 / NUMBER OF SEQ ID NOS: 23
 / SOFTWARE: PatentIn Ver. 2.0
 / SEQ ID NO 5
 / LENGTH: 428
 / TYPE: PRT
 / ORGANISM: Homo sapiens
 / US-10-091-085-5

Query Match 42.3%; Score 999; DB 13; Length 428;
 Best Local Similarity 52.4%; Pred. No. 1.4e-92;
 Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;
 QY 62 PLGTADGHEVYVYGMFDAGSTGRVHVFQPT-RPPREPTLTHETFAVKPGLSAYADD 120
 DB 40 PINVSA---STLYGMFDAGSTGRVHVFQPT-RPPREPTLTHETFAVKPGLSAYADD 96
 QY 121 VEKSAQIRELLDVAKODIPDFWKATPLVTKATAGTLLPGEKAKLLOKKEVFKASP 180
 DB 97 PRGAEVTVQGLLEVAKODIPDFWKATPLVTKATAGTLLPGEKAKLLOKKEVFKASP 156
 QY 181 FLVGDCCVIMNCTDEGSAMITINFLTGSILKTPGSSVGMGLDGGSTQIAFLPRVEGT 240
 DB 157 FLVPGKGSVIMKDSDEGLAWTVNVLGQLHGRQETVGTLLDGGASTQITFLPGEKTE 216
 QY 241 LQASPGVLTALMERNRYKYSYVGLGLMSARLALIGCVSGQPKADKELVSPCLSP 300
 DB 217 LEQTPRGVLTSEFEMENSTYKLYTHSYLGFGLKARLALIGCVSGQPKADKELVSPCLSP 275
 QY 301 SFKGEWENHAEVYTRVSGQAAASLHEICARVEVLONRVHRTVEVGVDFVAFSYYYDL 360
 DB 276 WLEAEWIVGVKYYQGGNQBEGVGFPEVYAEVLRVYQKHLHQBEPVRSFAFYFYDDAADTHL 335
 QY 361 AAGVGLDAEKGSILVGDPELAAYVCRLETOPQSSPSCMDLYVSLLOE-FGPPR 419
 DB 336 AVDTMIDYKGVLTSEFEMENSTYKLYTHSYLGFGLKARLALIGCVSGQPKADKELVSPCLSP 395

QY 420 SKVLKTRKIDNVTSWALGAFHYIDSL 448
Db 396 STVLQTKKVNITGKALGATFHLQSL 424

RESULT 12

US-10-092-063-3
; Sequence 3, Application US/10092063
; Publication No. US20020173005A1
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE POLYPEPTIDES
; FILE REFERENCE: 28110/35908
; CURRENT APPLICATION NUMBER: US/10/092,063
; CURRENT FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: 09/370,265
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350,836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-063-3

Query Match 42.3%; Score 999; DB 13; Length 428;

Best Local Similarity 52.4%; Pred. No. 1.4e-92;
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAADGHEVFYGMFDAGSTGTRVHVFOFT-RPPRETPTLTHTFFKAVKPGLSAYADD 120
Db 40 PINVSA---STLYGINFDAGSTGTRIHVYTFVQKMPGQLPILGEVDFSVFGLSAFVDQ 96
QY 121 VEKSAQIGIRELLDVAQDIPDFWKATPLVLKATAGRLLLPGEKAOKLLOKVEVFKASP 180
Db 97 PKQGAETVQGLLEAVKDSIPRSHWKKTPVVLKATAGRLLLPEHKAKALLFEVKEIFRKSP 156
QY 181 FLVGDDCVSTMGTDGVSAGWITINFLTGLSKATPGSSVGMLDLGGSGTQIAFLPRVEGT 240
Db 157 FLVPKGSVIMDSDEGILAWTVNFLTQQLHGHROETVGTLDLGGASTQITFLPQPEKT 216
QY 241 LQASPPGYLTALRMFNRTYKLSYSLGILMSARLAILGGVEGQPAKDGKELVSPCLSP 300
Db 217 LEQTPRGYLTSPFMFNSTVKLYTHSYLGFGLKAARLATLGALETE-GTDGHTFRSACLPR 275
QY 301 SPKGEWEHAEVTVRVSGQAAASLHELCAARYSEVLQNRVHRTVEVKHVDVFAFSYYIDL 360
Db 276 WLEAEWIFGKVYQYGGNOGEVGFPCYAEVLRVVRGKLHQPEEVQSGSFYAFSYYIDR 335
QY 361 AAGVGLIDAEKGGSLVVGDPFEIAAKVVCRTLETQPOSSPFCMDLTYVSLLIQOE-FGPPR 419
Db 336 AVDTDMIDYKGGILKVEDFERKAREVCNLENFTSGSPFLCMLDSYITALLKDGFGFAD 395
QY 420 SKVLKTRKIDNVTSWALGAFHYIDSL 448
Db 396 STVLQTKKVNITGKALGATFHLQSL 424

RESULT 13

US-10-092-063-5

; Sequence 5, Application US/10092063
; Publication No. US20020173005A1
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE POLYPEPTIDES
; FILE REFERENCE: 28110/35908
; CURRENT APPLICATION NUMBER: US/10/092,063
; CURRENT FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: 09/370,265
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350,836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-063-5

Query Match 42.3%; Score 999; DB 13; Length 428;

Best Local Similarity 52.4%; Pred. No. 1.4e-92;
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAADGHEVFYGMFDAGSTGTRVHVFOFT-RPPRETPTLTHTFFKAVKPGLSAYADD 120
Db 40 PINVSA---STLYGINFDAGSTGTRIHVYTFVQKMPGQLPILGEVDFSVKPGLSAFVDQ 96
QY 121 VEKSAQIGIRELLDVAQDIPDFWKATPLVLKATAGRLLLPGEKAOKLLOKVEVFKASP 180
Db 97 PKQGAETVQGLLEAVKDSIPRSHWKKTPVVLKATAGRLLLPEHKAKALLFEVKEIFRKSP 156
QY 181 FLVGDDCVSTMGTDGVSAGWITINFLTGLSKATPGSSVGMLDLGGSGTQIAFLPRVEGT 240
Db 157 FLVPKGSVIMDSDEGILAWTVNFLTQQLHGHROETVGTLDLGGASTQITFLPQPEKT 216
QY 241 LQASPPGYLTALRMFNRTYKLSYSLGILMSARLAILGGVEGQPAKDGKELVSPCLSP 300
Db 217 LEQTPRGYLTSPFMFNSTVKLYTHSYLGFGLKAARLATLGALETE-GTDGHTFRSACLPR 275
QY 301 SPKGEWEHAEVTVRVSGQAAASLHELCAARYSEVLQNRVHRTVEVKHVDVFAFSYYIDL 360
Db 276 WLEAEWIFGKVYQYGGNOGEVGFPCYAEVLRVVRGKLHQPEEVQSGSFYAFSYYIDR 335
QY 361 AAGVGLIDAEKGGSLVVGDPFEIAAKVVCRTLETQPOSSPFCMDLTYVSLLIQOE-FGPPR 419
Db 336 AVDTDMIDYKGGILKVEDFERKAREVCNLENFTSGSPFLCMLDSYITALLKDGFGFAD 395
QY 420 SKVLKTRKIDNVTSWALGAFHYIDSL 448
Db 396 STVLQTKKVNITGKALGATFHLQSL 424

RESULT 14

US-10-286-926-3

; Sequence 3, Application US/10286926
; Publication No. US20030175752A1
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; APPLICANT: Yeung, George
; TITLE OF INVENTION: Methods and Materials Relating to CD39-Like

Page 9

```

FILE REFERENCE: 28110/36457CON
CURRENT APPLICATION NUMBER: US/10/286,926
CURRENT FILING DATE: 2002-11-01
PRIOR APPLICATION NUMBER: 09/557,800
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/481,238
PRIOR FILING DATE: 2000-01-11
PRIOR APPLICATION NUMBER: 09/370,265
PRIOR FILING DATE: 1999-08-09
PRIOR APPLICATION NUMBER: PCT/US99/16180
PRIOR FILING DATE: 1999-07-16
PRIOR APPLICATION NUMBER: 09/350836
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: 09/273447
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: 09/122449
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: 09/244444
PRIOR FILING DATE: 1999-02-04
PRIOR APPLICATION NUMBER: 09/118,205
PRIOR FILING DATE: 1998-07-16
NUMBER OF SEQ ID NOS: 54
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 428
TYPE: PRT
ORGANISM: Homo sapiens
US-10-286-926-5

```

```

Best Local Similarity 52.4%;  Pred. No. 1,4e-92;
Matches 204;  Conservative 56;  Mismatches 123;  Indels 6;  Gaps 4;

QY      62  PLGTAADGHEVFQIMFDAGSTGRVAVFOFT-RPRPETFLTHETFAVKGSLAVAD 120
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      40  PINVSA---STVIGIMFDAGSTGRIRHIVTFVQKMPQQLPLBGEVFDVSKPGLSAFVD 96

QY      121  VERSAQGIRELLDVAKODIPDFEMKATPLVLKATAGRLILGEEKAKQLQKVEVFKASP 180
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      97  PKQGAETVQGLLEVAKXDSIPRSHMKTPVYLKATAGRLILPEHAKAKALLEVKEIFRKP 156

QY      181  FLVAGDCCVSIINNGTDEGVSAWITTNFLTGSLKTPGSGSSVGMLDLDGGSTQIAFLPRVEGT 240
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      157  FLVPKGSVSIMDGSDEGIILAMVTYNFLTGLHGRQETVGLDVGASTQITFLPQPEKT 216

QY      241  LQASPPGQITALLRMFNFTYKLISYSYIGLGMARRLILGVEGQAPAKDGKELVSPCLSP 300
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      217  LBQTFPRRYLTSFEHFNSTYKLYTHSYSGFGIKAKARLLTLGALETE-STDGHTFSACLPR 275

QY      301  SFKQEMEHAVTYTVRSQKAAASLHELCAARVSEVLONRVARTEEVHKVDVFARSSYYDL 360
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      276  WLBEMWTFGGVKVQYGANGGEVGFEBFCYAEVLKRVANGKLIHQPEVQSGSVFARSSYYDR 335

QY      361  AAGVGLIDAEKGSGLVGDPEILAKYCRILETOPGSSPFSCHDTLYVSLLLQE-FGRPR 419
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      336  AVDTMDIDYKRGGLKXKEDERRKAREVCNDLIENTSGSPFLCMDLSYITALLKDGFGPAD 395

QY      420  SKVLKLTFRKIDINVEVTSWALGAIFHYISL 448
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      396  STVLTQTKVNNIETGVALGATFTLLGSL 424

Search completed: June 8, 2004, 09:59:11
Job time : 51 secs

```

Tue Jun 8 12:10:38 2004

us-09-905-589a-2.ra1

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 09:48:55; Search time 22 Seconds
(without alignments)

1070.066 Million cell updates/sec

Title: US-09-905-589A-2

Perfect score: 2364
Sequence: 1 MRKISHHGSIRVAKVAVPLG.....ALGAIHYIDSLNRQKSPAS 456

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2364	100.0	456	US-09-240-639-2	Sequence 2, Appli
2	2364	100.0	484	US-09-608-285A-27	Sequence 27, Appli
3	2364	100.0	484	US-09-370-265-27	Sequence 27, Appli
4	2364	100.0	484	US-09-557-800C-27	Sequence 27, Appli
5	2364	100.0	484	US-09-370-625A-27	Sequence 27, Appli
6	2003	84.7	471	US-09-608-285A-60	Sequence 60, Appli
7	999	42.3	428	US-09-608-285A-3	Sequence 3, Appli
8	999	42.3	428	US-09-608-285A-5	Sequence 5, Appli
9	999	42.3	428	US-09-240-639-6	Sequence 6, Appli
10	999	42.3	428	US-09-240-639-9	Sequence 9, Appli
11	999	42.3	428	US-09-350-836B-3	Sequence 3, Appli
12	999	42.3	428	US-09-350-836B-5	Sequence 5, Appli
13	999	42.3	428	US-09-370-265-3	Sequence 3, Appli
14	999	42.3	428	US-09-370-265-5	Sequence 5, Appli
15	999	42.3	428	US-09-557-800C-3	Sequence 3, Appli
16	999	42.3	428	US-09-557-800C-5	Sequence 5, Appli
17	999	42.3	428	US-09-370-625A-3	Sequence 3, Appli
18	999	42.3	428	US-09-370-625A-5	Sequence 5, Appli
19	999	42.1	428	US-09-608-285A-7	Sequence 7, Appli
20	999	42.1	428	US-09-350-836B-7	Sequence 7, Appli
21	999	42.1	428	US-09-370-265-7	Sequence 7, Appli
22	999	42.1	428	US-09-557-800C-7	Sequence 7, Appli
23	999	42.1	428	US-09-370-625A-7	Sequence 7, Appli
24	909	38.5	405	US-09-608-285A-25	Sequence 25, Appli
25	909	38.5	405	US-09-370-265-25	Sequence 25, Appli
26	909	38.5	405	US-09-557-800C-25	Sequence 25, Appli
27	909	38.5	405	US-09-370-625A-25	Sequence 25, Appli

28	909	38.5	465	US-09-557-800C-56	Sequence 56, Appli
29	909	38.5	465	US-09-370-625A-39	Sequence 39, Appli
30	904	38.2	465	US-09-240-639-8	Sequence 8, Appli
31	525.5	22.2	467	US-09-129-112-19	Sequence 19, Appli
32	506	21.4	462	US-09-129-112-2	Sequence 2, Appli
33	501	21.2	459	US-09-129-112-9	Sequence 9, Appli
34	485	20.5	454	US-09-240-639-11	Sequence 11, Appli
35	466	19.7	473	US-09-240-639-12	Sequence 12, Appli
36	463	19.6	462	US-09-129-112-15	Sequence 15, Appli
37	442.5	18.7	455	US-09-240-639-10	Sequence 10, Appli
38	314.5	13.3	529	US-09-240-639-4	Sequence 4, Appli
39	285.5	12.1	502	US-09-557-800C-55	Sequence 55, Appli
40	285.5	12.1	502	US-09-370-625A-38	Sequence 38, Appli
41	285.5	12.1	510	US-08-930-921-1	Sequence 1, Appli
42	216.5	9.2	148	US-09-240-639-17	Sequence 17, Appli
43	184	7.8	150	US-09-240-639-16	Sequence 16, Appli
44	183.5	7.8	154	US-09-240-639-14	Sequence 14, Appli
45	175.5	7.4	153	US-09-240-639-15	Sequence 15, Appli

ALIGNMENTS

```
RESULT 1
US-09-240-639-2
; Sequence 2, Application US/09240639
; Patent No. 635047
; GENERAL INFORMATION:
; APPLICANT: Chadwick, Brian Paul
; APPLICANT: Frischau, Anna-Maria
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO CD39-LIKE
; TITLE OF INVENTION: POLYPEPTIDES AND NUCLEIC ACIDS
; FILE REFERENCE: 9598-066
; CURRENT APPLICATION NUMBER: US/09/240, 639
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-240-639-2
```

Query Match 100.0%; Score 2364; DB 4; Length 456;
Best Local Similarity 100.0%; Pred. No. 5.8e+240;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MRKISHHGSIRVAKVAVPLGICVGVFIYVYKMRATATQAFPSITRAAPGARWQOAH	60
DB	1	MRKISHHGSIRVAKVAVPLGICVGVFIYVYKMRATATQAFPSITRAAPGARWQOAH	60
QY	61	SPGTAADGHEVYVYGMFAGSTGTGVHVFQFTRPREPTLTHETFKAVKGLSAVDD	120
DB	61	SPGTAADGHEVYVYGMFAGSTGTGVHVFQFTRPREPTLTHETFKAVKGLSAVDD	120
QY	121	VERSAQIGHEILDVAQDIPDFWKATPLVLAATAGLRLLPEBKAKQLQKYEYFKASP	180
DB	121	VERSAQIGHEILDVAQDIPDFWKATPLVLAATAGLRLLPEBKAKQLQKYEYFKASP	180
QY	181	FLVGGDCVSTIMNGTDBGSAMTINLTGSLTPGGSSGMDLGGGSGTQIAFLPVEET	240
DB	181	FLVGGDCVSTIMNGTDBGSAMTINLTGSLTPGGSSGMDLGGGSGTQIAFLPVEET	240
QY	241	LQASPPGYLTALMFNRYTKLYSYSLGLGMSARLAIIGVYEQAPADKELVSPCLSP	300
DB	241	LQASPPGYLTALMFNRYTKLYSYSLGLGMSARLAIIGVYEQAPADKELVSPCLSP	300
QY	301	SFGGEWEHAEVYVYVGGQRAASLHELCAARSEVYQNVHRTTEKGVDFVAFSYVDL	360
DB	301	SFGGEWEHAEVYVYVGGQRAASLHELCAARSEVYQNVHRTTEKGVDFVAFSYVDL	360
QY	361	AAGVGILDEKGGSLVVGDFEIAKTYVCTLETOPSSPSCMDLTVYSLILQETGPPRS	420
DB	361	AAGVGILDEKGGSLVVGDFEIAKTYVCTLETOPSSPSCMDLTVYSLILQETGPPRS	420

Db 361 AAGVGLIDAEKGGSLVVGDFEIAAKVCRTELETPQSSPFCMDLTYSLLQLQEGFPFRS 420
Qy 421 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456
Db 421 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456
RESULT 2
US-09-608-285A-27
; Sequence 27, Application US/09608285A
; Patent No. 6335013
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Muleto, George
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE
; FILE REFERENCE: 28110/36570
; CURRENT FILING DATE: 2000-06-30
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-04-25
; PRIOR FILING DATE: 2000-04-25
; PRIOR FILING DATE: 2000-01-11
; PRIOR FILING DATE: 2000-01-11
; PRIOR FILING DATE: 1999-08-09
; PRIOR FILING DATE: 1999-08-09
; PRIOR FILING DATE: 1999-07-16
; PRIOR FILING DATE: 1999-07-16
; PRIOR FILING DATE: 1999-07-09
; PRIOR FILING DATE: 1999-07-09
; PRIOR FILING DATE: 1999-03-19
; PRIOR FILING DATE: 1999-03-19
; PRIOR FILING DATE: 1999-02-04
; PRIOR FILING DATE: 1999-02-04
; PRIOR FILING DATE: 1998-07-24
; PRIOR FILING DATE: 1998-07-24
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 27
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-608-285A-27
Query Match 100.0%; Score 2364; DB 4; Length 484;
Best Local Similarity 100.0%; Pred. No. 6.4e-240;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFPSITRAAPGARWQQA 60
Db 29 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFPSITRAAPGARWQQA 88
Qy 61 SPLGTAADGHEVYGINFMDAGSTGTRVHVQFTRPPRETPTLTHETFKAVKPLSAYADD 120
Db 89 SPLGTAADGHEVYGINFMDAGSTGTRVHVQFTRPPRETPTLTHETFKAVKPLSAYADD 148
Qy 121 VEKSAQIGIRELLDVAKODIPDFWKATPLVKATAGLRLLPGEKAQKLLQKVEFKASP 180
Db 149 VEKSAQIGIRELLDVAKODIPDFWKATPLVKATAGLRLLPGEKAQKLLQKVEFKASP 208
Qy 181 FLVGDDCVSINMGTDGVSAMITINFLTGLSKTPGGSSVGMGLDGGSGTQIAFLPRVEGT 240
Db 209 FLVGDDCVSINMGTDGVSAMITINFLTGLSKTPGGSSVGMGLDGGSGTQIAFLPRVEGT 268
Qy 241 LOASPPGYLTALRMFNNTYKLYSYSLGLGMSARLAILGGVEGQPAKDGKELVSPCLSP 300
Db 269 FLVGDDCVSINMGTDGVSAMITINFLTGLSKTPGGSSVGMGLDGGSGTQIAFLPRVEGT 328
Qy 301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEEVKVDVFAFSYYIDL 360
Db 361 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEEVKVDVFAFSYYIDL 360

Db 329 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEEVKVDVFAFSYYIDL 388
Qy 361 AAGVGLIDAEKGGSLVVGDFEIAAKVCRTELETPQSSPFCMDLTYSLLQLQEGFPFRS 420
Db 389 AAGVGLIDAEKGGSLVVGDFEIAAKVCRTELETPQSSPFCMDLTYSLLQLQEGFPFRS 448
Qy 421 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456
Db 449 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 484
RESULT 3
US-09-370-265-27
; Sequence 27, Application US/09370265
; Patent No. 6447771
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Muleto, Julio
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
; FILE REFERENCE: 28111/35908
; CURRENT FILING DATE: 1999-08-09
; PRIOR FILING DATE: 1999-08-09
; PRIOR FILING DATE: 1999-07-16
; PRIOR FILING DATE: 1999-07-16
; PRIOR FILING DATE: 1999-07-09
; PRIOR FILING DATE: 1999-07-09
; PRIOR FILING DATE: 1999-03-19
; PRIOR FILING DATE: 1999-03-19
; PRIOR FILING DATE: 1999-02-04
; PRIOR FILING DATE: 1999-02-04
; PRIOR FILING DATE: 1998-07-24
; PRIOR FILING DATE: 1998-07-24
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 27
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-370-265-27
Query Match 100.0%; Score 2364; DB 4; Length 484;
Best Local Similarity 100.0%; Pred. No. 6.4e-240;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFPSITRAAPGARWQQA 60
Db 29 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFPSITRAAPGARWQQA 88
Qy 61 SPLGTAADGHEVYGINFMDAGSTGTRVHVQFTRPPRETPTLTHETFKAVKPLSAYADD 120
Db 89 SPLGTAADGHEVYGINFMDAGSTGTRVHVQFTRPPRETPTLTHETFKAVKPLSAYADD 148
Qy 121 VEKSAQIGIRELLDVAKODIPDFWKATPLVKATAGLRLLPGEKAQKLLQKVEFKASP 180
Db 149 VEKSAQIGIRELLDVAKODIPDFWKATPLVKATAGLRLLPGEKAQKLLQKVEFKASP 208
Qy 181 FLVGDDCVSINMGTDGVSAMITINFLTGLSKTPGGSSVGMGLDGGSGTQIAFLPRVEGT 240
Db 209 FLVGDDCVSINMGTDGVSAMITINFLTGLSKTPGGSSVGMGLDGGSGTQIAFLPRVEGT 268
Qy 241 LOASPPGYLTALRMFNNTYKLYSYSLGLGMSARLAILGGVEGQPAKDGKELVSPCLSP 300
Db 269 LOASPPGYLTALRMFNNTYKLYSYSLGLGMSARLAILGGVEGQPAKDGKELVSPCLSP 328
Qy 301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEEVKVDVFAFSYYIDL 360
Db 329 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEEVKVDVFAFSYYIDL 388
Qy 361 AAGVGLIDAEKGGSLVVGDFEIAAKVCRTELETPQSSPFCMDLTYSLLQLQEGFPFRS 420

Db 389 AAGVGLIDAEKGSGLVGVDFEIAKYVCRTLETPOSSPSCMDLTYVSLLOEFGFPRS 448
Qy 421 KVLKTRKIDNVETSWALGAIFFHYIDSLNRQKSPAS 456
Db 449 KVLKTRKIDNVETSWALGAIFFHYIDSLNRQKSPAS 484

RESULT 4

US-09-557-800C-27
Sequence 27, Application US/09557800C
Patent No. 6476211
GENERAL INFORMATION:
APPLICANT: Ford, John
APPLICANT: Mulero, Julio
APPLICANT: Yeung, George
TITLE OF INVENTION: Methods and Materials Relating to CD39-Like
FILE REFERENCE: 28110/36457
CURRENT APPLICATION NUMBER: US/09/557,800C
CURRENT FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/461,238
PRIOR FILING DATE: 2000-01-11
PRIOR APPLICATION NUMBER: 09/370,265
PRIOR FILING DATE: 1999-08-09
PRIOR APPLICATION NUMBER: PCT/US99/16180
PRIOR FILING DATE: 1999-07-16
PRIOR APPLICATION NUMBER: 09/350836
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: 09/273447
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: 09/122449
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: 09/244444
PRIOR FILING DATE: 1999-02-04
PRIOR APPLICATION NUMBER: 09/118,205
PRIOR FILING DATE: 1998-07-16
NUMBER OF SEQ ID NOS: 56
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 27
LENGTH: 484
TYPE: PRT
ORGANISM: Homo sapiens
US-09-557-800C-27

Query Match 100.0%; Score 2364; DB 4; Length 484;
Best Local Similarity 100.0%; Pred. No. 6 4e-240;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRKISNHSGLRVAKAVPLGLCVGVFTYVAYIKMHRATATQAFPSITRAAPGARWGOAH 60
Db 29 MRKISNHSGLRVAKAVPLGLCVGVFTYVAYIKMHRATATQAFPSITRAAPGARWGOAH 88
Qy 61 SPLGTADGHEVFGIMFDAGSTGRVHVFOFTPRPPTPTLTHTETKAVKPGLSAYAD 120
Db 89 SPLGTADGHEVFGIMFDAGSTGRVHVFOFTPRPPTPTLTHTETKAVKPGLSAYAD 148
Qy 121 VEKSAQGIREDLDVAKODIPDFWKATPVLKATAGRLPGEKAQKLLQKVEVFKA 180
Db 149 VEKSAQGIREDLDVAKODIPDFWKATPVLKATAGRLPGEKAQKLLQKVEVFKA 208
Qy 181 FLVDDCVSINMGDEGVSAWITTNFLTGLSKTPGSSVGMULDGGSTQIAFLPVEGT 240
Db 209 FLVDDCVSINMGDEGVSAWITTNFLTGLSKTPGSSVGMULDGGSTQIAFLPVEGT 268
Qy 241 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLILGVVEGQPAKDGKEIVSPCLSP 300
Db 269 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLILGVVEGQPAKDGKEIVSPCLSP 328
Qy 301 SFKGEWEAEVTVYVSGQKAAASLHELCAARVSEVLQNRVHRTVEVGHVDFYAFSYYDL 360
Db 329 SFKGEWEAEVTVYVSGQKAAASLHELCAARVSEVLQNRVHRTVEVGHVDFYAFSYYDL 388

Qy 361 AAGVGLIDAEKGSGLVGVDFEIAKYVCRTLETPOSSPSCMDLTYVSLLOEFGFPRS 420
Db 389 AAGVGLIDAEKGSGLVGVDFEIAKYVCRTLETPOSSPSCMDLTYVSLLOEFGFPRS 448
Qy 421 KVLKTRKIDNVETSWALGAIFFHYIDSLNRQKSPAS 456
Db 449 KVLKTRKIDNVETSWALGAIFFHYIDSLNRQKSPAS 484

RESULT 5

US-09-370-625A-27
Sequence 27, Application US/09370625A
Patent No. 660032
GENERAL INFORMATION:
APPLICANT: Ford, John
APPLICANT: Mulero, Julio
APPLICANT: Yeung, George
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE POLYPEPTIDES
FILE REFERENCE: 28110/35908
CURRENT APPLICATION NUMBER: US/09/370,625A
CURRENT FILING DATE: 1999-08-09
PRIOR APPLICATION NUMBER: PCT/US99/16180
PRIOR FILING DATE: 1999-07-16
PRIOR APPLICATION NUMBER: 09/350,836
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: 09/273,447
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 27
LENGTH: 484
TYPE: PRT
ORGANISM: Homo sapiens
US-09-370-625A-27

Query Match 100.0%; Score 2364; DB 4; Length 484;
Best Local Similarity 100.0%; Pred. No. 6 4e-240;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRKISNHSGLRVAKAVPLGLCVGVFTYVAYIKMHRATATQAFPSITRAAPGARWGOAH 60
Db 29 MRKISNHSGLRVAKAVPLGLCVGVFTYVAYIKMHRATATQAFPSITRAAPGARWGOAH 88
Qy 61 SPLGTADGHEVFGIMFDAGSTGRVHVFOFTPRPPTPTLTHTETKAVKPGLSAYAD 120
Db 89 SPLGTADGHEVFGIMFDAGSTGRVHVFOFTPRPPTPTLTHTETKAVKPGLSAYAD 148
Qy 121 VEKSAQGIREDLDVAKODIPDFWKATPVLKATAGRLPGEKAQKLLQKVEVFKA 180
Db 149 VEKSAQGIREDLDVAKODIPDFWKATPVLKATAGRLPGEKAQKLLQKVEVFKA 208
Qy 181 FLVDDCVSINMGDEGVSAWITTNFLTGLSKTPGSSVGMULDGGSTQIAFLPVEGT 240
Db 209 FLVDDCVSINMGDEGVSAWITTNFLTGLSKTPGSSVGMULDGGSTQIAFLPVEGT 268
Qy 241 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLILGVVEGQPAKDGKEIVSPCLSP 300
Db 269 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLILGVVEGQPAKDGKEIVSPCLSP 328
Qy 301 SFKGEWEAEVTVYVSGQKAAASLHELCAARVSEVLQNRVHRTVEVGHVDFYAFSYYDL 360
Db 329 SFKGEWEAEVTVYVSGQKAAASLHELCAARVSEVLQNRVHRTVEVGHVDFYAFSYYDL 388
Qy 361 AAGVGLIDAEKGSGLVGVDFEIAKYVCRTLETPOSSPSCMDLTYVSLLOEFGFPRS 420
Db 389 AAGVGLIDAEKGSGLVGVDFEIAKYVCRTLETPOSSPSCMDLTYVSLLOEFGFPRS 448
Qy 421 KVLKTRKIDNVETSWALGAIFFHYIDSLNRQKSPAS 456
Db 449 KVLKTRKIDNVETSWALGAIFFHYIDSLNRQKSPAS 484

RESULT 6
US-09-608-285A-60

; Sequence 60, Application US/09608285A
; Patent No. 6335013
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; APPLICANT: Yeung, George
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: 28110/36570
; CURRENT APPLICATION NUMBER: US/09/608,285A
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/583,231
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 09/557,800
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/481,238
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 09/370,265
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350,836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-608-285A-60

Query Match 84.7%; Score 2003; DB 4; Length 471;
Best Local Similarity 98.5%; Pred. No. 6.3e-202;
Matches 388; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
Qy 1 MRKISNHGSLRVAKVAYPLGCVGVFIYVAYIKWHRATATQAFPSITRAAPGARWQQAH 60
Db 29 MRKISNHGSLRVAKVAYPLGCVGVFIYVAYIKWHRATATQAFPSITRAAPGARWQQAH 88
Qy 61 SPLGTAADGHEVFGIMFDAGSTGTRVHVFOFTPRPREPTTLTHTFKAVKPLSAYADD 120
Db 89 SPLGTAADGHEVFGIMFDAGSTGTRVHVFOFTPRPREPTTLTHTFKAVKPLSAYADD 148
Qy 121 VEKSAQGIREDLDVAKQDIPDFWKATPLVLKATAGRLLLPGEKAKLLQKVKEVFKASP 180
Db 149 VEKSAQGIREDLDVAKQDIPDFWKATPLVLKATAGRLLLPGEKAKLLQKVKEVFKASP 208
Qy 181 FLVGDDCVSIMGTDGVSAMITINFLTGSLKTPGSSVGMGLDGGSTQIAFLPRVEGT 240
Db 209 FLVGDDCVSIMGTDGVSAMITINFLTGSLKTPGSSVGMGLDGGSTQIAFLPRVEGT 268
Qy 241 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 300
Db 269 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 328
Qy 301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTEEVKHVDYFAFSYYDL 360
Db 329 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTEEVKHVDYFAFSYYDL 388
Qy 361 AAGVGLIDAEKGGSLVVGDFEIAAKVCRTELEQ 394
Db 389 AAGVGLIDAEKGGSLVVGDFEIAAKVCRTELEQ 422

RESULT 7

US-09-608-285A-3
; Sequence 3, Application US/09608285A
; Patent No. 6335013
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; APPLICANT: Yeung, George
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: 28110/36570
; CURRENT APPLICATION NUMBER: US/09/608,285A
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/583,231
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 09/557,800
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/481,238
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 09/370,265
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350,836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-608-285A-3
Query Match 42.3%; Score 999; DB 4; Length 428;
Best Local Similarity 52.4%; Pred. No. 2.8e-96;
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;
Qy 62 PLGTAADGHEVFGIMFDAGSTGTRVHVFOFTPRPREPTTLTHTFKAVKPLSAYADD 120
Db 40 PINVSA--STLYGIMFDAGSTGTRVHVFOFTGQMPQQLPILGEVFDSEVKGPLSAFVDQ 96
Qy 121 VEKSAQGIREDLDVAKQDIPDFWKATPLVLKATAGRLLLPGEKAKLLQKVKEVFKASP 180
Db 97 PKGAETVQGLLELVAKDIPRSHWKTFFVVLKATAGRLLLPGEKAKLLQKVKEVFKASP 156
Qy 181 FLVGDDCVSIMGTDGVSAMITINFLTGSLKTPGSSVGMGLDGGSTQIAFLPRVEGT 240
Db 157 FLVPGSVSIMGSDGILAWTVNFLTQHLGHRQETVGTDLGGASTQITLPPQPEKT 216
Qy 241 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 300
Db 217 LEQTPRGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 275
Qy 301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTEEVKHVDYFAFSYYDL 360
Db 276 WLEAEWIFGGYKQYGGQGEVGFPCYAEVLVRVGRKLHQPEEVQSGFYAFSYYDR 335
Qy 361 AAGVGLIDAEKGGSLVVGDFEIAAKVCRTELEQ 394
Db 336 AVDTMDIDYKGGSLVVGDFEIAAKVCRTELEQ 395
Qy 420 SKVLKLRKIDNVETSWALGAIFHYIDSL 448
Db 396 STVQLQIKKVNNTGHWALGATFHLLQSL 424


```
Query Match 42.3%; Score 999; DB 4; Length 428;
Best Local Similarity 52.4%; Pred. No. 2.8e-96;
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAADGHEVYFYGIMFDAGSTGRVHVFOFT-RPPRETTLTHETFKAVKPGLSAYADD 120
DB 40 PINVSA---STLYGIMFDAGSTGRVHVFOFT-RPPRETTLTHETFKAVKPGLSAYADD 96

QY 121 VEKSAQGIREDLLVAKODIPDFWKATPLVLKATAGLRLLPGEKAOKLLOKVEFKASP 180
DB 97 PKQGAETVQGLLEVAKODIPDFWKATPLVLKATAGLRLLPGEKAOKLLOKVEFKASP 156

QY 181 FLVGDDCVSINMGTDGVSAMITINFLTGLSGKTPGSSVGMGLDGGSGTOIAPLPRVEGT 240
DB 157 FLVPGKGSVIMDGSDGILAWTVNFLTGLGHRQETVGTDLGGASTQITFLPQFEKT 216

QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGSGVEGQPAKDGKELVSPCLSP 300
DB 217 LEQTPRGYLTSPFNFNSTYKLYTHSYLGLGMSARLAILGSGVEGQPAKDGKELVSPCLSP 275

QY 301 SFKGEWEHAEVTVYRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVKHVDYAFSYYDL 360
DB 276 WLEAEWIFGGVYKYQYGGNQEVEGFPCYAEVLVVRGKLHQPVEVQSGSYAFSYYDL 335

QY 361 AAGVGLIDAEKGGSLVGDGFIAAKYVCRITLQPOSSPSCMDLTYSVLLLOE-FGPPR 419
DB 336 AVDTMDIDYKGGILKVEDFERKAREVCNLENFTSGSPFLCMLDLYTALLKDGFGFAD 395

QY 420 SKVLKTRKIDNVTSMALGAIPHVIDSL 448
DB 396 STVLQTLTKVNNIETGALGATFHLQSL 424

RESULT 11
US-09-350-836B-3
; Sequence 3, Application US/09350836B
; Patent No. 6387645
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
; FILE REFERENCE: 28110/35761
; CURRENT APPLICATION NUMBER: US/09/350,836B
; CURRENT FILING DATE: 1999-07-09
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-350-836B-3

Query Match 42.3%; Score 999; DB 4; Length 428;
Best Local Similarity 52.4%; Pred. No. 2.8e-96;
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAADGHEVYFYGIMFDAGSTGRVHVFOFT-RPPRETTLTHETFKAVKPGLSAYADD 120
DB 40 PINVSA---STLYGIMFDAGSTGRVHVFOFT-RPPRETTLTHETFKAVKPGLSAYADD 96

QY 121 VEKSAQGIREDLLVAKODIPDFWKATPLVLKATAGLRLLPGEKAOKLLOKVEFKASP 180
DB 97 PKQGAETVQGLLEVAKODIPDFWKATPLVLKATAGLRLLPGEKAOKLLOKVEFKASP 156

QY 181 FLVGDDCVSINMGTDGVSAMITINFLTGLSGKTPGSSVGMGLDGGSGTOIAPLPRVEGT 240
DB 157 FLVPGKGSVIMDGSDGILAWTVNFLTGLGHRQETVGTDLGGASTQITFLPQFEKT 216

QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGSGVEGQPAKDGKELVSPCLSP 300
DB 217 LEQTPRGYLTSPFNFNSTYKLYTHSYLGLGMSARLAILGSGVEGQPAKDGKELVSPCLSP 275

QY 301 SFKGEWEHAEVTVYRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVKHVDYAFSYYDL 360
DB 276 WLEAEWIFGGVYKYQYGGNQEVEGFPCYAEVLVVRGKLHQPVEVQSGSYAFSYYDL 335

QY 361 AAGVGLIDAEKGGSLVGDGFIAAKYVCRITLQPOSSPSCMDLTYSVLLLOE-FGPPR 419
DB 336 AVDTMDIDYKGGILKVEDFERKAREVCNLENFTSGSPFLCMLDLYTALLKDGFGFAD 395

QY 420 SKVLKTRKIDNVTSMALGAIPHVIDSL 448
DB 396 STVLQTLTKVNNIETGALGATFHLQSL 424

RESULT 12
US-09-350-836B-5
; Sequence 5, Application US/09350836B
; Patent No. 6387645
; GENERAL INFORMATION:
; APPLICANT: Mulero, Julio
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
; FILE REFERENCE: 28110/35761
; CURRENT APPLICATION NUMBER: US/09/350,836B
; CURRENT FILING DATE: 1999-07-09
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-350-836B-5
```

```
Query Match 42.3%; Score 999; DB 4; Length 428;
Best Local Similarity 52.4%; Pred. No. 2.8e-96;
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAADGHEVYFYGIMFDAGSTGRVHVFOFT-RPPRETTLTHETFKAVKPGLSAYADD 120
DB 40 PINVSA---STLYGIMFDAGSTGRVHVFOFT-RPPRETTLTHETFKAVKPGLSAYADD 96

QY 121 VEKSAQGIREDLLVAKODIPDFWKATPLVLKATAGLRLLPGEKAOKLLOKVEFKASP 180
DB 97 PKQGAETVQGLLEVAKODIPDFWKATPLVLKATAGLRLLPGEKAOKLLOKVEFKASP 156

QY 181 FLVGDDCVSINMGTDGVSAMITINFLTGLSGKTPGSSVGMGLDGGSGTOIAPLPRVEGT 240
DB 157 FLVPGKGSVIMDGSDGILAWTVNFLTGLGHRQETVGTDLGGASTQITFLPQFEKT 216

QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGSGVEGQPAKDGKELVSPCLSP 300
DB 217 LEQTPRGYLTSPFNFNSTYKLYTHSYLGLGMSARLAILGSGVEGQPAKDGKELVSPCLSP 275

QY 301 SFKGEWEHAEVTVYRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVKHVDYAFSYYDL 360
DB 276 WLEAEWIFGGVYKYQYGGNQEVEGFPCYAEVLVVRGKLHQPVEVQSGSYAFSYYDL 335

QY 361 AAGVGLIDAEKGGSLVGDGFIAAKYVCRITLQPOSSPSCMDLTYSVLLLOE-FGPPR 419
DB 336 AVDTMDIDYKGGILKVEDFERKAREVCNLENFTSGSPFLCMLDLYTALLKDGFGFAD 395

QY 420 SKVLKTRKIDNVTSMALGAIPHVIDSL 448
DB 396 STVLQTLTKVNNIETGALGATFHLQSL 424

RESULT 12
US-09-350-836B-5
; Sequence 5, Application US/09350836B
; Patent No. 6387645
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
; FILE REFERENCE: 28110/35761
; CURRENT APPLICATION NUMBER: US/09/350,836B
; CURRENT FILING DATE: 1999-07-09
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-350-836B-5
```

QY 361 AAGVGLDAEKGGSLVVGDFEIAKVCRTLETQPOSSPCMDLTYVSLLOE-FGPR 419
Db 336 AVDTMDIDYKGGILKVEDFERKAREVCNLENFTSGSPFLCMDLSYITALLKDGFGFAD 395
QY 420 SKVLKTRKIDNVTSMALGAI FHYIDSL 448
Db 396 STVLQLTTRKVNNIETGMALGATFHLQSL 424

RESULT 13

US-09-370-265-3
Sequence 3, Application US/09370265
Patent No. 6447771
GENERAL INFORMATION:
APPLICANT: Mulero, John
APPLICANT: Mulero, Jullio
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
POLYPEPTIDES
FILE REFERENCE: 2811/35908
CURRENT APPLICATION NUMBER: US/09/370,265
EARLIER FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: PCT/US99/16180
EARLIER FILING DATE: 1999-07-16
EARLIER APPLICATION NUMBER: 09/350,836
EARLIER FILING DATE: 1999-07-09
EARLIER APPLICATION NUMBER: 09/273,447
EARLIER FILING DATE: 1999-03-19
EARLIER APPLICATION NUMBER: 09/244,444
EARLIER FILING DATE: 1999-02-04
EARLIER APPLICATION NUMBER: 09/122,449
EARLIER FILING DATE: 1998-07-24
EARLIER APPLICATION NUMBER: 09/118,205
EARLIER FILING DATE: 1998-07-16
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 428
TYPE: PRT
ORGANISM: Homo sapiens
US-09-370-265-3

Query Match 42.3%; Score 999; DB 4; Length 428;
Best Local Similarity 52.4%; Pred. No. 2.8e-96;
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAAADGHEVYFVGMFDAGSTGRVHVFOFT-RPRRETPTLTHETFKAVKPGLSAVYD 120
Db 40 PINVSA---STLYGMFDAGSTGRVHVFOFT-RPRRETPTLTHETFKAVKPGLSAVYD 96
QY 121 VEKSAQGIREFLDVAKODIPDFWKATPLVKATAGRLLPGEKAKQLLOKVEYFKASP 180
Db 97 PKQAEVYQGLLEVAKDISPSHMKKTVPVLKATAGRLLPGEKAKQLLOKVEYFKASP 156
QY 181 FLVGGDDCVSINMGDEGVSANITINFLTGSUKTGGSSVGMULDGGSTQIAFLPRVEGT 240
Db 157 FLVPGKSVSINMGDEGVSANITINFLTGSUKTGGSSVGMULDGGSTQIAFLPRVEGT 216
QY 241 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGVEGQPADGKELVSPCLSP 300
Db 217 LEQTPRGVLTSPFNENSTYKLYSYSLGLGMSARLAILGVEGQPADGKELVSPCLSP 275
QY 301 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTVEYKVDFAESYYVDL 360
Db 276 WLEAEWIFGVYKYQYGVNOGEVGEPCYAEVLRVVRKGLQHPBEVORGSEFYAFSYYDR 335
QY 361 AAGVGLDAEKGGSLVVGDFEIAKVCRTLETQPOSSPCMDLTYVSLLOE-FGPR 419
Db 336 AVDTMDIDYKGGILKVEDFERKAREVCNLENFTSGSPFLCMDLSYITALLKDGFGFAD 395
QY 420 SKVLKTRKIDNVTSMALGAI FHYIDSL 448
Db 396 STVLQLTTRKVNNIETGMALGATFHLQSL 424

RESULT 14

US-09-370-265-5
Sequence 5, Application US/09370265
Patent No. 6447771
GENERAL INFORMATION:
APPLICANT: Ford, John
APPLICANT: Mulero, Jullio
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
POLYPEPTIDES
FILE REFERENCE: 2811/35908
CURRENT APPLICATION NUMBER: US/09/370,265
EARLIER FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: PCT/US99/16180
EARLIER FILING DATE: 1999-07-16
EARLIER APPLICATION NUMBER: 09/350,836
EARLIER FILING DATE: 1999-07-09
EARLIER APPLICATION NUMBER: 09/273,447
EARLIER FILING DATE: 1999-03-19
EARLIER APPLICATION NUMBER: 09/244,444
EARLIER FILING DATE: 1999-02-04
EARLIER APPLICATION NUMBER: 09/122,449
EARLIER FILING DATE: 1998-07-24
EARLIER APPLICATION NUMBER: 09/118,205
EARLIER FILING DATE: 1998-07-16
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 428
TYPE: PRT
ORGANISM: Homo sapiens
US-09-370-265-5

Query Match 42.3%; Score 999; DB 4; Length 428;
Best Local Similarity 52.4%; Pred. No. 2.8e-96;
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAAADGHEVYFVGMFDAGSTGRVHVFOFT-RPRRETPTLTHETFKAVKPGLSAVYD 120
Db 40 PINVSA---STLYGMFDAGSTGRVHVFOFT-RPRRETPTLTHETFKAVKPGLSAVYD 96
QY 121 VEKSAQGIREFLDVAKODIPDFWKATPLVKATAGRLLPGEKAKQLLOKVEYFKASP 180
Db 97 PKQAEVYQGLLEVAKDISPSHMKKTVPVLKATAGRLLPGEKAKQLLOKVEYFKASP 156
QY 181 FLVGGDDCVSINMGDEGVSANITINFLTGSUKTGGSSVGMULDGGSTQIAFLPRVEGT 240
Db 157 FLVPGKSVSINMGDEGVSANITINFLTGSUKTGGSSVGMULDGGSTQIAFLPRVEGT 216
QY 241 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGVEGQPADGKELVSPCLSP 300
Db 217 LEQTPRGVLTSPFNENSTYKLYSYSLGLGMSARLAILGVEGQPADGKELVSPCLSP 275
QY 301 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTVEYKVDFAESYYVDL 360
Db 276 WLEAEWIFGVYKYQYGVNOGEVGEPCYAEVLRVVRKGLQHPBEVORGSEFYAFSYYDR 335
QY 361 AAGVGLDAEKGGSLVVGDFEIAKVCRTLETQPOSSPCMDLTYVSLLOE-FGPR 419
Db 336 AVDTMDIDYKGGILKVEDFERKAREVCNLENFTSGSPFLCMDLSYITALLKDGFGFAD 395
QY 420 SKVLKTRKIDNVTSMALGAI FHYIDSL 448
Db 396 STVLQLTTRKVNNIETGMALGATFHLQSL 424

RESULT 15

US-09-557-800C-3
Sequence 3, Application US/09557800C
Patent No. 6476211
GENERAL INFORMATION:
APPLICANT: Ford, John

APPLICANT: Mulero, Julio
APPLICANT: Yeung, George
TITLE OF INVENTION: Methods and Materials Relating to CD39-Like
FILE REFERENCE: 28110/36457
CURRENT APPLICATION NUMBER: US/09/557,800C
CURRENT FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/481,238
PRIOR FILING DATE: 2000-01-11
PRIOR APPLICATION NUMBER: 09/370,265
PRIOR FILING DATE: 1999-08-09
PRIOR APPLICATION NUMBER: PCT/US99/16180
PRIOR FILING DATE: 1999-07-16
PRIOR APPLICATION NUMBER: 09/350836
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: 09/273447
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: 09/122449
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: 09/244444
PRIOR FILING DATE: 1999-02-04
PRIOR APPLICATION NUMBER: 09/118,205
PRIOR FILING DATE: 1998-07-16
NUMBER OF SEQ ID NOS: 56
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 428
TYPE: PRT
ORGANISM: Homo sapiens
US-09-557-800C-3

Query Match 42.3%; Score 999; DB 4; Length 428;
Best Local Similarity 52.4%; Pred. No. 2.8e-96;
Matches 204; Conservative 56; Mismatches 123; Indels 5; Gaps 4;

Qy	62	PLGTADGHEVFYGMFDAGSTGTRVHVFOFT-RPPREPTTLTHETFKAVKPLGSLAYADD	120
Db	40	PINVSA---STLYGIMFDAGSTGTRVHVFOFTGQMFQQLFLEGEVDSVKPGLSAFVDQ	96
Qy	121	VEKSAQGIREDLVAKODIPDFWKATPLVLKATAGLRLLPGEKAKLOKVKVEFKASP	180
Db	97	PKQAEVQGLLEVAKDSIPRSHWKKTPVVLKATAGLRLLPEKAKALLFEVKEIFRKSP	156
Qy	181	FLVGDDCVSIMNGTDEGVSAWITINFLTGLSKTPGSSVGMLDLGGSGTOIAFLPRVEGT	240
Db	157	FLVPKGSVIMDGSDEGILAWTVNFLTQGLHGHQROETVGTLDLGGASTQITFLPQFEKT	216
Qy	241	LQASPPGYLTALSMFNETHYKLYSYSLGILMSARLAILGGVEGQPAKDGKELVSPCLSP	300
Db	217	LEOTPRGYLTSEFMFNSTYKLYTHSYGLFGLKAARLATLGALETE-GTDGHTFRSACLPR	275
Qy	301	SFKGEWEHAEVTVRVSCQKAAASLHELCAARVSEVLQNRVHRTVEEVKHVDYFAFSYYDL	360
Db	276	WLEAEWIFGGVKYQYGGNQGSEVGFPCYAEVLRVVRGKLHQFEEVQSGSFYAFSYYDR	335
Qy	361	AAGVLIDAEKGSLLVVGDPFIAAKVYCRTELETPQOSSPFCMDLTYVSLILQE-FGPPR	419
Db	336	AVDTDMIDYEKGGILKVEDFERKAREVCNLENFTSGSPFLCMDLSYITALLKDGFGFAD	395
Qy	420	SKVLKLTFRKIDNVETSWALCAIFHYIDSL	448
Db	396	STVLQITKVNNIETGWALGATFHLLQSL	424

Search completed: June 8, 2004, 09:57:36
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 09:58:16 ; Search time 20 Seconds

(without alignments)
577.149 Million cell updates/sec

Title: US-09-905-589a-2_COPY_1_120

Perfect score: 640
Sequence: 1 MKKSNHGSRLKAKVAVPLG.....TLTHETPKAVKPGLSAVDD 120

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 52131

Minimum DB seq length: 0
Maximum DB seq length: 120

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	61	9.5	99 2 S70668	chaperonin groES -
2	56	8.8	115 2 G72466	hypothetical prote
3	53.5	8.4	110 2 S58804	ubiquinol-cytochro
4	53.5	8.4	113 2 AH0397	probable lipoprote
5	53	8.3	113 2 T09627	positive transcrip
6	52.5	8.2	117 2 A87029	hypothetical prote
7	51.5	8.0	79 2 G95324	probable TrpA con
8	51.5	8.0	110 2 S58803	ubiquinol-cytochro
9	51	8.0	81 2 A25903	T-cell receptor be
10	50.5	7.9	100 2 PQ0115	hypothetical prote
11	50.5	7.9	103 2 F70909	hypothetical prote
12	50	7.8	116 2 T13211	minor capsid prote
13	49.5	7.7	106 2 S38768	rho protein GDP di
14	49	7.7	38 2 T06971	hypothetical prote
15	49	7.7	101 2 GMHUB	gastrin precursor
16	49	7.7	113 2 AG0573	transcription acti
17	48.5	7.6	91 1 H63198	hypothetical prote
18	48.5	7.6	106 2 AE2371	hypothetical prote
19	48.5	7.6	107 1 W4MLB4	E4 protein - bovin
20	48.5	7.6	111 2 A87525	BF hand domain pro
21	48	7.5	68 2 E82779	hypothetical prote
22	48	7.5	83 2 T07257	cytochrome b559 co
23	48	7.5	107 2 T70537	hypothetical prote
24	48	7.5	111 2 S45396	Ig lambda chain V
25	47.5	7.4	84 1 WMA5PW	gene p protein - p
26	47.5	7.4	109 2 D75538	conserved hypotet
27	47.5	7.4	112 2 S26221	cruciferin (clone
28	47.5	7.4	114 2 F48364	hypothetical prote
29	47.5	7.4	115 2 S62526	hypothetical prote

30	47	7.3	41 2 F84020	hypothetical prote
31	47	7.3	82 2 E84290	hypothetical prote
32	47	7.3	84 2 T33867	hypothetical prote
33	47	7.3	102 2 C83315	NADH2 dehydrogenas
34	47	7.3	102 2 H87678	hypothetical prote
35	47	7.3	106 2 B64759	membrane protein y
36	47	7.3	116 2 S36004	NADH2 dehydrogenas
37	46.5	7.3	72 2 T17855	hypothetical prote
38	46.5	7.3	97 2 S07266	hypothetical prote
39	46.5	7.3	102 2 T34786	hypothetical prote
40	46.5	7.3	106 2 T14819	dehydrin DHN2 - so
41	46.5	7.3	106 2 S61101	hypothetical prote
42	46.5	7.3	118 1 R7MCM1	ribosomal protein
43	46	7.2	76 2 S28250	NADH2 dehydrogenas
44	46	7.2	90 2 H86964	hypothetical prote
45	46	7.2	95 2 H69341	conserved hypotet

ALIGNMENTS

RESULT 1
S70668
chaperonin groES - Caulobacter crescentus
N/Alternate names: heat shock protein groES
C/Species: Caulobacter crescentus
C/Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C/Accession: S70668; S70666
R/Avedissian, M.; Gomes, S. J.
submitted to the EMBL Data Library, April 1995
A/Description: Expression of the groESL operon is cell cycle controlled in Caulobacter
A/Reference number: S70668
A/Accession: S70668
A/Molecule type: DNA
A/Residues: 1-99 <AVE>
A/Cross-references: EMBL:J41394; NID:G1663717; PIDN:AAB18634.1; PID:G763440
A/Experimental source: strain NA1000
R/Avedissian, M.; Lopes Gomes, S.
Mol. Microbiol. 19, 79-89, 1996
A/Title: Expression of the groESL operon is cell-cycle controlled in Caulobacter crescentus
A/Reference number: S70666; WUID:96419165; PMID:8621538
A/Accession: S70666
A/Molecule type: DNA
A/Residues: 1-18 <AV2>
A/Cross-references: EMBL:J41394
A/Experimental source: strain NA1000
C/Genetics:
A/Gene: groES
C/Superfamily: chaperonin groES
C/Keywords: ATP; heat shock; molecular chaperone; stress-induced protein

Query Match 9.5%; Score 61; DB 2; Length 99;
Best Local Similarity 36.4%; Pred. No. 15;
Matches 16; Conservative 6; Mismatches 20; Indels 2; Gaps 2;

QY 46 ITPAPCARWGQAHSPPLGTADGHEVFQIMFDAGSTGRVHV 89
Db 40 VVRSGPGAR-NHETSSPLDVKA-GDRILFGKMGSGTASEGTXYKV 81

RESULT 2
G72466
hypothetical protein APE2378 - Aeropyrum pernix (strain K1)
C/Species: Aeropyrum pernix
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C/Accession: G72466
R/Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999
A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum
A/Reference number: A72450; WUID:99310339; PMID:10382966
A/Accession: G72466
A/Status: preliminary

[illegible]

Matches	19; Conservative	10; Mismatches	26; Indels	21; Gaps	4;
QY	3	KINSHGSLRYAKVAYYPPLGLCVGFIVYAYIKMRHATATQAFSTTPAAGSARWGQAHSP	62		
Db	47	RVDGHSRVLVTMKSLL-----IRVAAY-WRRPAL-----AYFLPQRESRRDP	89		
QY	63	L-----GTRADGHEVFY	74		
Db	90	AMTMDETVASHGHTFW	105		

RESULT 7
G59324
Probable Trad conjugal transfer protein [imported] - *Sinorhizobium meliloti* (strain 1021)
C:Species: *Sinorhizobium meliloti*
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: G95324
R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bower, J.; Kalmann, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888 2001
A:Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*
A:Reference number: A95262; MUID:21396509; PMID:11481452
A:Accession: G95324
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-79 <KUR>
A:Cross-references: GS:AE006469; PIDN:PAK65161.1; PID:G14523604; GSPDB:GN00165
A:Experimental source: Strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, J.; Chait, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L. Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalmann, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, hebaull, P.; Vandenbol, M.; Vortroler, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.K.
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; MUID:21368234; PMID:11474104
C:Contents: annotation
C:Genetics:
A:Gene: trad
A:Genome: plasmid

Query Match	8.0%	Score 51.5	DB 2	Length 79
Best Local Similarity	27.4%	Pred. No. 1.3e+02		
Matches 20	Conservative 6	Mismatches 38	Indels 9	Gaps 3

QY 13 AKVAFLGLCVGVFIYAYIKMHRATADAFSSITRAAPARKGOAHSPLGTALDGHV 72
| : || : | | : | : |
Db 9 AREKFLG---AIVVRAGLTKADRAFLLGGFIELARVTPDS---AEHRRLDI--GEAA 59

QY	73	FYGMFDAGSTGT	85
Db	60	FKAPALDGGSPGT	72

RESULT 8
558803
ubiquinol-cytochrome-c reductase (BC 1.10.2.2) cytochrome b - *Doronomymex kutleri* mitoc
C|Species: mitochondrion *Doronomymex kutleri*
C|Date: 19-Mar-1997 #sequence_revision 24-Oct-1997 #text_change 03-Jun-2002
C|Accession: 558803
R|Baur, A.; Chalvatzis, N.; Buschinger, A.; Zimmermann, F.K.
C|Date: Genet. 28, 242-247, 1995
A|Title: Mitochondrial DNA sequences reveal close relationships between social parasitoid
A|Reference number: 558798; MIMD:96071193; PMID:8525270
A|Accession: 558803
A|Molecule type: DNA
A|Residues: 1-110 <BAU>
C|Genetics:
A|Genome: mitochondrion
C|Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastoquinol
C|Keywords: chromoprotein; electron transfer; heme; iron; mitochondrion; oxidative phosph
F|1-110/Domain: cytochrome b homology (fragment) <CB6>
F|1-110/Domain: cytochrome b homology (fragment) <CBH>

```
F:9-25/Domain: transmembrane #status predicted <TM1>
F:54-72/Domain: transmembrane #status predicted <TM2>
F:90-106/Domain: transmembrane #status predicted <TM3>
```

Query Match	5.0%	Score	51.5	DB 2	Length	110	
Best Local Similarity	29.3%	Pred. No.	1	8e+02			
Matches	24	Conservative	10	Mismatches	35	Gaps	4

QY 5 SNHSLRVAKVAYYPGLCVGVFIYVA-YIKMRATATQ-AFFSLTRAAIGARWGOAHP 62

DB 4 SNYGSU-----LGLFLAIQIFSGEFLSIHYCPNTLAFSLIHIMONISNGMMHN- 54

```

QY      63 LGTAADGHEVFYGYIMFDAGSTG 84
      55 --IHNGASFFFCMYIHMSRG 74
Db

```

```

RESULT 9
A:Accession: A25903
C:Cell receptor beta chain precursor V region (Mlp2) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 23-Jul-1999
C:Accession: A25903
R:Stinger, P.A.; McEvilly, R.J.; Noonan, D.J.; Dixon, F.J.; Theofilopoulos, A.N.
Proc. Natl. Acad. Sci. U.S.A. 83, 7018-7022, 1986
A:Title: Clonal diversity in T-cell receptor beta-chain variable gene expression in
A:Reference number: A25903; MUID:86313702; PMID:3462739
A:Accession: A25903
A:Molecule type: mRNA
A:Residues: 1-81 <SB>
A:Cross-references: GI:M14294; NID:g201349; PIDN:AAA0241.1; PID:g201350
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: T-cell receptor
F:1-62/Domain: V segment (fragment) <VAR>
F:63-84/Domain: D segment <DIV>
F:85-81/Domain: J segment <JO1>

```

Query Match	8.0%;	Score 51;	DB 2;	Length 81;
Similarity	29.2%;	Pred. No. 1.5e+02;		
Matches	19;	Conservative	24;	Indels 10; Gaps 3

QY 52 GAATGCGAAGSPLCTGAADGHEVFVGIMFPAAGSTGRVHVFQTRPRPREFTLTETFKAVK 111
|||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 2 GAATGCGALMTATANEGSEAT---ESGFTDK---PPISRNLTSTL---TWNAR 51

Qy	112	PGLSA	116
			:
Db	52	PGDS	56

RESULT 10
PQ0115

RESULT 10
 PQ0115
 hypochlorite protein (azu region) - *Pseudomonas aeruginosa* (fragment)
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 31-Mar-1992 #sequence_revision 31-Dec-1993 #text_change 08-Oct-1999
 C:Accession: PQ0115
 R:HotLink, C.W.G.; Woudt, L.P.; Threnhout, J.C.M.; van de Kamp, M.; Canters, G.W.
 Gene 90, 15-20, 1990
 A:Title: Isolation and sequencing of the *Alcaligenes denitrificans* azurin-encoding gene
 A:Reference number: JQ0643; MUID:90337337; PMID:2116366
 A:Accession: PQ0115
 A:Molecule type: DNA
 A:Residues: 1-100 <HOT>
 A:Cross-references: GB:930389, NID:g151060, PIDN:AAA25731.1, PID:g151063

Query Match	7.9%	Score 50.5;	DB 2;	Length 100;
Best Local Similarity	34.2%	Pred. 50.2;	1e+02;	
Matches 13; Conservative	4;	Mismatches 16;	Indels 5;	Gaps 1;

QY	22	CACVETIYAYIKWH-----RATATQAFPSITRAAPGAR	54
Db	62	CANPDQYFWDEWHPTREHVHOLGEAMAAATREFGAR	99

C;Accession: S38768
R;Pick, E.; Gorzalczany, Y.; Engel, S.
Eur. J. Biochem. 217, 441-455, 1993

A;Title: Role of the rac1 p21-GDP-dissociation inhibitor for rho heterodimer in the actin
A;Reference number: S38767; MUID:94039069; PMID:8223583

A;Accession: S38768
A:Molecule type: protein
A;Residues: 1-40;41-46;47-86;87-106 <PIC>
C;Superfamily: human rho GDP dissociation inhibitor

Query Match 7.7%; Score 49.5; DB 2; Length 106;
Best Local Similarity 29.3%; Pred. No. 2.9e+02;
Matches 12; Conservative 4; Mismatches 20; Indels 5; Gaps 1;

QY 70 HEVFGIMED-----AGSGTGVTHVFQFTRPPRETPTLTHE 105
DB HTKRGKVGIKDIDYMVGSYGPRAEYEFLLTFMEAPKFDD 90

RESULT 14
T06971
hypothetical protein 32 - Cyanophora paradoxa cyanelle
C;Species: cyanelle Cyanophora paradoxa
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 12-Jun-2003
C;Accession: T06971
R;Stirrewalt, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohnert, H.J.; Bryant, D.A.
submitted to the EMBL Data Library, July 1995
A;Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.
A;Reference number: Z15840
A;Accession: T06971
A;Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A;Residues: 1-38 <STI>
A;Cross-references: EMBL:U30821; NID:g1016083; PIDN:AAA81314.1; PID:g1016227
A;Experimental source: strain Pringsheim LB555
C;Genetics:
A;Gene: ycf32
A;Genome: cyanelle
C;Superfamily: uncharacterized conserved protein ycf32
C;Keywords: cyanelle

Query Match 7.7%; Score 49; DB 2; Length 38;
Best Local Similarity 34.4%; Pred. No. 1.1e+02;
Matches 11; Conservative 7; Mismatches 10; Indels 4; Gaps 1;

QY 9 SLRVAKVAVPLGLCVGVFIY---VAYIKWHR 36
DB 2 SNELVVVLPLGTALGWVYNICKLAIEQWR 33

RESULT 15
GMHUB
gastrin precursor [validated] - human
N;Contains: big gastrin; cryptagastrin; gastrin; gastrin-17
C;Species: Homo sapiens (man)
C;Date: 24-Apr-1984 #sequence_revision 30-Jun-1987 #text_change 19-Jan-2001
C;Accession: A93997; A93497; A94473; A93152; A91628; A18854; A40869; A32487; B32487; C32
R;Ito, R.; Sato, K.; Helmer, T.; Jay, G.; Agarwal, K.
Proc. Natl. Acad. Sci. U.S.A. 81, 4662-4666, 1984
A;Title: Structural analysis of the gene encoding human gastrin: the large intron contain
A;Reference number: A93997; MUID:84272693; PMID:6087340
A;Accession: A93997
A:Molecule type: DNA
A;Residues: 1-101 <TO>
A;Cross-references: GB:K01254; GB:J00147; NID:g182987; PIDN:AAB59533.1; PID:g182989
R;Kato, K.; Hayashiaki, Y.; Takahashi, Y.; Himeno, S.; Matsubara, K.
Nucleic Acids Res. 11, 8197-8203, 1983
A;Title: Molecular cloning of the human gastrin gene.
A;Reference number: A93497; MUID:84169471; PMID:6324077
A;Accession: A93497
A:Molecule type: DNA
A;Residues: 1-101 <XAT>
A;Cross-references: GB:X00183; NID:g31648; PIDN:CAA25005.1; PID:g31649

R.Harris, U.I.; Kenner, E.W.
 unpublished results, cited by Gregory, R.A., and Tracy, H.J., in Gastrointestinal Hormon
 A:Reference number: A94473
 A:Accession: A94473
 A:Molecule type: protein
 A:Residues: 59-64, 'H', '66', 'S', '68-92 <HAR>
 A:Experimental source: Gastrinoma tissue
 R.Bentley, P.H.; Kenner, G.W.; Sheppard, R.C.
 Nature 209, 583-585, 1966
 A:Title: Human gastrin isolation, structure and synthesis.
 A:Reference number: A93152; PMID:67021327; PMID:5921183
 A:Accession: A93152
 A:Molecule type: protein
 A:Residues: 76-92 <BEN>
 A:Note: gastrin comprises the carboxyl-terminal 17 residues of big gastrin
 R.Gregory, R.A.; Tracy, H.J.; Agarwal, K.L.; Grossman, M.I.
 Gut 10, 603-608, 1969
 A:Title: Amino acid constitution of two gastrins isolated from Zollinger-Ellison tumour
 A:Reference number: A91628; PMID:69298172; PMID:5822140
 A:Accession: A91628
 A:Molecule type: protein
 A:Residues: 76-92 <GRE>
 A:Note: the gastrin was isolated from human Zollinger-Ellison pancreatic tumor
 R.Wiborg, O.; Berglund, L.; Boel, E.; Norris, P.; Norris, K.; Rehfeld, J.F.; Marcker, K.
 Proc. Natl. Acad. Sci. U.S.A. 81, 1067-1069, 1984
 A:Title: Structure of a human gastrin gene.
 A:Reference number: A18854; PMID:84144842; PMID:6322186
 A:Accession: A18854
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-101 <MIB>
 A:Cross-references: GB:K01254; GB:U00147; NID:G182967; PTDN:AA59533.1; PTD:G182989
 R.Hubner, V.D.; Jiang, R.; Lee, T.D.; Legesse, K.; Walsh, J.H.; Shively, J.E.; Chew, P.
 J. Biol. Chem. 266, 12223-12227, 1991
 A:Title: Purification and structural characterization of progastrin-derived peptides frd
 A:Reference number: A40869; PMID:91286536; PMID:2061307
 A:Accession: A40869
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 22-51 <HUB>
 R.Hiashimoto, Y.; Himeno, S.; Shinomura, Y.; Nagao, K.; Tamura, T.; Tani, S.
 Biochem. Biophys. Res. Commun. 160, 1364-1370, 1989
 A:Title: Purification and structural determination of urinary NH-2-terminal big gastrin
 A:Reference number: A32487; PMID:89273602; PMID:2730647
 A:Accession: A32487
 A:Molecule type: protein
 A:Residues: 59-67 <HI2>
 A:Note: this urinary fragment of big gastrin was designated peak Ia
 A:Experimental source: urine
 A:Accession: C32487
 A:Molecule type: protein
 A:Residues: 59-66 <HI3>
 A:Note: this urinary fragment of big gastrin was designated peak Ib
 A:Experimental source: urine
 A:Accession: C32487
 A:Molecule type: protein
 A:Residues: 59-68 <HI4>
 A:Note: this urinary fragment of big gastrin was designated peak II
 R.Hiashimoto, Y.; Shinomura, Y.; Nagao, K.; Yasunaga, Y.; Tamura, T.; Tani, S.
 Biochem. Biophys. Res. Commun. 172, 1392-1399, 1990
 A:Title: Purification of N-terminal hexapeptide of big gastrin from human urine.
 A:Reference number: A36249; PMID:91058586; PMID:2244919
 A:Accession: A36249
 A:Molecule type: protein
 A:Residues: 59-64 <HI5>
 A:Note: this urinary fragment of big gastrin was designated peak III
 R.Boel, E.; Vunatt, J.; Norris, P.; Norris, K.; Wind, A.; Rehfeld, J.F.; Marcker, K.A.
 Proc. Natl. Acad. Sci. U.S.A. 80, 2866-2869, 1983
 A:Title: Molecular cloning of human gastrin cDNA: evidence for evolution of gastrin by g
 A:Reference number: 137408; PMID:83321503; PMID:6574456
 A:Accession: 137408
 A:Molecule type: mRNA

A:Residues: 1-101 <RES>
 A:Cross-references: EMBL:V00511; NID:G31654; PTDN:CAA23769.1; PTD:G31655
 R.Rehfeld, J.F.; Johnsen, A.H.
 Eur. J. Biochem. 223, 765-773, 1994
 A:Title: Identification of gastrin component I as gastrin-71. The largest possible bloc
 A:Reference number: S48183; PMID:9433379; PMID:8053952
 A:Accession: S48183
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 22-40 <REH>
 R.Kariya, Y.; Kato, K.; Hayashizaki, Y.; Himeno, S.; Tani, S.; Matsubara, K.
 Gene 50, 345-352, 1986
 A:Title: Expression of human gastrin gene in normal and gastrinoma tissues.
 A:Reference number: 154006; PMID:87219893; PMID:3034736
 A:Accession: 154006
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-101 <KAR>
 A:Cross-references: GB:M15958; NID:G182990; PTDN:AAA52520.1; PTD:G182991
 R.Rehfeld, J.F.; Hansen, C.P.; Johnsen, A.H.
 EMBO J. 14, 389-396, 1995
 A:Title: Post-poly(Glu) cleavage and degradation modified by O-sulfated tyrosine: a nove
 A:Reference number: S54350; PMID:95137019; PMID:7530658
 A:Accession: S54350
 A:Molecule type: protein
 A:Residues: 76-92 <KEM>
 A:Note: gastrin-6 isolated from gastric mucosa and blood is fully active; sulfation of 1
 C:Genetics:
 A:Gene: GDB:GAS
 A:Cross-references: GDB:119261; OMIM:137250
 A:Map position: 17q-17q
 A:Functions: 71/1
 C:Superfamily: gastrin
 C:Keywords: amidated carboxyl end; phosphoprotein; pyroglutamic acid; sulfoxprotein
 F:1/21/Domain: signal sequence #status predicted <SIG>
 F:2/25/Product: cryptogastria (amino-terminal propeptide) #status experimental <PRO>
 F:59-92/Product: big gastrin #status experimental <BGN>
 F:76-92/Product: gastrin #status experimental <SGN>
 F:87-92/Product: gastrin-6 #status experimental <GN6>
 F:55/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #link BGN #status
 F:76/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #link SGN #status
 F:87/Binding site: sulfate (Tyr) (covalent) (partial) #status experimental
 F:92/Modified site: amidated carboxyl end (Phe) (amide in mature form from following Gly
 Query Match 7.7%; Score 49; DB 1; Length 101;
 Best Local Similarity 35.3%; Pred. No. 3.1e+02;
 Matches 18; Conservative 3; Mismatches 12; Indels 18; Gaps 3;
 QY 21 LCYGVFIYVAYIKMHRATATCAFFSITRAAPARW---GQAHSPILGTAAD 68
 Db 4 LCYVVLIF-----ALAAAFSE-----ASWKPRSQQDAPLGTGAN 39

Search completed: June 8, 2004, 10:02:12
 Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 09:57:10 ; Search time 17 Seconds
(without alignments)

367.554 Million cell updates/sec

Title: US-09-905-589A-2_COPY_1_120

Perfect score: 640
Sequence: 1 MRKISNHSGLRVAKVAVPLG.....TLHETPKAVKPLGLSAVAD 120

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 20892

Minimum DB seq length: 0

Maximum DB seq length: 120

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	8.3	113	1	RAMA_KLEPN
2	52	8.1	113	1	RAMA_KLEPN
3	51	8.0	101	1	PER2_MOUSE
4	49.5	7.7	117	1	PAND_BRAVA
5	49	7.7	38	1	PSBY_CYAPA
6	49	7.7	101	1	GAST_HUMAN
7	48.5	7.6	107	1	VEA_BVVA
8	48.5	7.6	111	1	GDIF_CAVPO
9	48	7.5	83	1	PSBE_CHLVV
10	47.5	7.4	83	1	VPP_BPRD
11	47.5	7.4	115	1	VPP_BPRD
12	47.5	7.4	115	1	YC9C_SCHPO
13	47	7.3	106	1	YAHM_ECOLI
14	47	7.3	116	1	NUJM_CYPCA
15	46.5	7.3	97	1	YI09_EMENT
16	46.5	7.3	106	1	YNE3_YEAST
17	46.5	7.3	118	1	RL7_MICLU
18	46	7.2	76	1	NIKV_BOVIN
19	46	7.2	93	1	Y736_ARCFU
20	45.5	7.1	65	1	YPH2_SYNP2
21	45.5	7.1	102	1	TRAC_RHISN
22	45.5	7.1	106	1	ONCB_TAERA
23	45	7.0	83	1	PSBE_NEPOL
24	45	7.0	101	1	PER2_HUMAN
25	45	7.0	109	1	YDUD_BACSV
26	45	7.0	116	1	BRJ1_BRARE
27	44.5	7.0	85	1	CRH_BACSV
28	44.5	7.0	103	1	FER_SULAC
29	44.5	7.0	113	1	Y956_CAMTE
30	44	6.9	67	1	CERC_CERCA
31	44	6.9	82	1	PSBE_CHIRE
32	44	6.9	94	1	Y518_METUA
33	44	6.9	111	1	LV2H_HUMAN

ALIGNMENTS

34	44	6.9	111	1	YHB2_STROO
35	44	6.9	116	1	PO41_CHICK
36	44	6.9	118	1	4YB1_HUMAN
37	43.5	6.8	67	1	YVBI_VACCC
38	43.5	6.8	80	1	R31E_STRMU
39	43.5	6.8	82	1	P2Y2_MERIN
40	43.5	6.8	98	1	ETOX_RHIME
41	43.5	6.8	109	1	PT16_STYPL
42	43.5	6.8	111	1	Y5ND_BACSV
43	43.5	6.8	117	1	DHA_CRAPL
44	43.5	6.8	117	1	HV10_MOUSE
45	43.5	6.8	117	1	RBPV_LACPL
					P23159 streptomyc
					P55968 gallus gall
					Q13541 homo sapien
					P20549 vaccinia vi
					Q84tn5 streptococc
					Q99pe3 meriones un
					Q02730 rhizobium m
					P28208 etyela plic
					P94563 bacillus su
					P22238 craterostig
					P01754 mus musculu
					Q88vk8 lactobacill

RESULT 1

AC	Q48413	STANDARD	PRT	113 AA.
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Transcriptional activator rama.			
GN	RAMA.			
OS	Klebsiella pneumoniae.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Klebsiella.			
OX	NCBI_TaxID=573;			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=EC18;			
RX	MEDLINE=96032015; PubMed=7551053;			
RA	George A.M., Hall R.M., Stokes H.W.;			
RT	"Multidrug resistance in Klebsiella pneumoniae: a novel gene, rama,			
RT	confers a multidrug resistance phenotype in Escherichia coli.";			
RL	Microbiology 141:1909-1920(1995).			
CC	-1- FUNCTION: PROBABLE TRANSCRIPTIONAL ACTIVATOR. CONFERS A MULTIDRUG			
CC	RESISTANCE PHENOTYPE IN ESCHERICHIA COLI.			
CC	-1- SIMILARITY: BELONGS TO THE ARAC/XYLIS FAMILY OF TRANSCRIPTIONAL			
CC	REGULATORS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U19581; AAA85697.1; -			
DR	PIR; T03627; T03627.			
DR	HSSP; P27246; 1BL0.			
DR	InterPro; IPR000005; HTHARAC.			
DR	Pfam; PF00165; HTH_ARAC; 2.			
DR	PRINTS; PR00032; HTHARAC.			
DR	SMART; SM00342; HTH_ARAC; 1.			
DR	PROSITE; PS00041; HTH_ARAC FAMILY 1; 1.			
DR	PROSITE; PS01124; HTH_ARAC FAMILY 2; 1.			
KW	Transcription regulation; DNA-binding; Activator.			
FT	DNA BIND 25 44 H-T-H motif (POTENTIAL).			
SC	SEQUENCE 113 AA; 13495 MW; 46E0D415C6C6C9F CRC64;			
QY	Query Match 8.3%; Score 53; DB 1; Length 113;			
QY	Best Local Similarity 26.0%; Pred. No. 68;			
QY	Matches 25; Conservative 9; Mismatches 42; Indels 20; Gaps 4;			
DB	7 HGSIRAKYAVPLGCVGIFIVAYIKYRAATQAFSITMAAGARWGQQAHSPLG- 65			
DB	20 HGPLRIDDIARHNG-----YSKW---LQRLFYQYGSISIGRYIRERKULLAAR 65			
QY	66 -AADGHEVFYGM---FDAGSTGRVHVFCFTRPP 96			

Db	66	DLRDTQRYVDYDCLKYGFSQTFVFTFTFNQPP	101	113	AA.
RESULT 2					
RAMA	ENTCL				
ID	RAMA	ENTCL	STANDARD;	PRT;	113 AA.
AC	P5922;				
DT	01-NOV-1997	(Rel. 35, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DE	15-JUL-1998	(Rel. 36, Last annotation update)			
DE	Transcriptional activator	ramA.			
OS	Enterobacteriaceae				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Enterobacter.				
OX	NCBI_TaxID=550;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=90299838; PubMed=2193928;				
RA	Komatsu T., Ohta M., Kido N., Akakawa Y., Ito H., Mizuno T., Kato N.;				
RT	"Molecular characterization of an Enterobacter cloacae gene (romA)				
RT	which pleiotropically inhibits the expression of Escherichia coli				
RT	outer membrane proteins.";				
RL	J. Bacteriol. 172:4082-4089 (1990).				
RN	[2]				
RP	IDENTIFICATION.				
RX	MEDLINE=96032015; PubMed=7551053;				
RA	George A.M., Hall R.M., Stokes H.W.;				
RT	"Multidrug resistance in Klebsiella pneumoniae: a novel gene, ramA,				
RT	confers a multidrug resistance phenotype in Escherichia coli.";				
RL	Microbiology 141:1909-1920 (1995).				
CC	-1- FUNCTION: Probable transcriptional activator.				
CC	-1- SIMILARITY: BELONGS TO THE ARAC/XYLS FAMILY OF TRANSCRIPTIONAL				
CC	REGULATORS.				
DR	HSP; P27246; 1BL0.				
DR	InterPro: IPR000005; HTHArac.				
DR	Pfam: PF00163; HTH_Arac; 2.				
DR	PRINTS: PM00032; HTHARAC.				
DR	SMART: SM00342; HTH_ARAC; 1.				
DR	PROSITE; PS00041; HTH_ARAC FAMILY 1; FALSE NEG.				
DR	PROSITE; PS01124; HTH_ARAC FAMILY 2; 1.				
KW	Transcription regulation; DNA-binding; Activator.				
FT	DNA BIND 25 44				
FT	H-T-H MOTIF (POTENTIAL).				
SQ	SEQUENCE 113 AA; 13633 MW; D9D57E77D64B9557 CRC64;				
Query Match	8.1%; Score 52; DB 1; Length 113;				
Best Local Similarity	24.1%; Pred. No. 88;				
Matches	26; Conservative 13; Mismatches 45; Indels 24; Gaps 5;				
Qy	7 HGSRLVAKVPLGLCVGVFIYVAYIKVHRATATQAFSITRAAPGARGQQAHSLPGLT- 65				
Db	20 HQPLRIEDIAHAG-----YSKWH-----LQRLFLQYKGSGLRYIREKLLLAAR 65				
Qy	66 -AAGCHEVFGIM----FDAGSTGTRVHVVFQTRPP-----RETFTLTH 104				
Db	66 DLRSDDRVVEICLRYGFSQQTFTFTFTTQHPGAYRKENHSRTH 113				
RESULT 3					
PEN2 MOUSE					
ID	PEN2 MOUSE	STANDARD;	PRT;	101 AA.	
AC	Q9C0R7;				
DT	10-OCT-2003	(Rel. 42, Created)			
DT	10-OCT-2003	(Rel. 42, Last sequence update)			
DT	10-OCT-2003	(Rel. 42, Last annotation update)			
GN	Gamma-secretase subunit PEN-2 (Presenilin enhancer protein 2).				
GN	PEN2.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				

```

DR EMBL AK007621: BAB25141.1; -.
DR EMBL BC024347: AAH24347.1; -.
DR MGI:1913590: 1700023M9Rik.
KM Transmembrane; Golgi stack; Endoplasmic reticulum.
FT DOMAIN 1 17 LUMENAL (POTENTIAL).
FT TRANSMEM 18 38 POTENTIAL.
FT DOMAIN 39 60 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 61 81 POTENTIAL.
FT DOMAIN 82 101 LUMENAL (POTENTIAL).
SQ SEQUENCE 101 AA; 11999 MW; 6596CA2984FBC59 CRC64;

Query Match 8.0%; Score 51; DB 1; Length 101;
Best Local Similarity 23.2%; Pred. No. 1e+02;
Matches 22; Conservative 15; Mismatches 44; Indels 14; Gaps 3;

QY 1 MKETNSHSLRVAKVAPLGLGVGVIVVATIKHRAATQAFFSITRAAGAWGQAH 60
DB 3 LERVNNEKINICRKYIGGFAPLPLVNVIFW-----FREAFLAPAYTEGQIK 54
QY 61 SPLGTADGHEVFGIMFDAGSTGTRVHVFQTRP 95
DB 55 GYVMSAVGF-LFWIIL-----ATWITIFQIYRP 83

RESULT 4
PAND_BRAJA STANDARD; PRT; 117 AA.
ID PAND_BRAJA
AC Q9AMG6;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 10-OCT-2003 (Rel. 42; Last annotation update)
DE Aspartate 1-decarboxylase precursor (EC 4.1.1.11) (Aspartate alpha-decarboxylase).
OS Brachyrrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobium.
ON NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110SPC4;
RX MEDLINE=21101824; PubMed=1157954;
RA Goettfert M., Roethlisberger S., Kuendig C., Beck C., Marty R., Henecke H.;
RT "Potential symbiosis-specific genes uncovered by sequencing a 410-kb DNA region of the Bradyrhizobium japonicum chromosome.";
RL J. Bacteriol. 183:1405-1412 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=2248498; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamitawa K., Uchiyama T., Sasamoto S., Ideasa K., Iritani Y., Kawashima K., Kohara M., Matsunoto M., Shimpo S., Tsunoda H., Wada T., Yamada M., Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium Brachyrrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197 (2002).
RN [3]
RP CARBOLYTIC ACTIVITY: L-aspartate = beta-alanine + CO(2).
CC -1- COFACTOR: Pyruvoyl group (by similarity).
CC -1- PATHWAY: Pantothenate biosynthesis; second branch.
CC -1- SIMILARITY: Belongs to the pand family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.ebi.ac.uk/announcements/ or send an email to license@ebi.ac.uk).
CC EMBL AF323013; AAG1071.1; -.
CC EMBL AP005942; BAC47361.1; -.

```

```

DR HSSP; P31664; 1AMB.
DR HAMAP; MF 00446; -1.
DR InterPro; IPR009010; Asp_decarb_fold.
DR InterPro; IPR003150; Asp_decarb.
DR Pfam; PF0261; Asp_decarb.1.
DR ProDom; PD009294; Asp_decarb.1.
DR TIGRFAMs; TIGR00223; pand; 1.
KM Pantothenate biosynthesis; Lyase; Decarboxylase; Pyruvate; Zymogen; Complete proteome.
FT CHAIN 1 24 ASPARTATE 1-DECARBOXYLASE BETA CHAIN (BY SIMILARITY).
FT FT ASPARTATE 1-DECARBOXYLASE ALPHA CHAIN (BY SIMILARITY).
FT MOD_RES 25 25 CONVERTED TO A PYRUVYL GROUP (BY SIMILARITY).
SQ SEQUENCE 117 AA; 12895 MW; B468238A20B50RCE CRC64;

Query Match 7.7%; Score 49.5; DB 1; Length 117;
Best Local Similarity 26.8%; Pred. No. 1.7e+02;
Matches 26; Conservative 8; Mismatches 38; Indels 25; Gaps 5;

QY 33 KMRATATQA-----FSTRAAPGAPGQQAHSPLGTADGHEVFGIMFDAGSTGR 86
DB 9 KIRASVTADLHYBSISIDRTLEA-----AGVIVREIVNV-----ETGR 54
QY 87 VHVQFTPPRE-TPTLHETFRVAVPG---LSAVA 118
DB 55 FAVYIEAPPMSGTMSLNGAARLVWPGDITIVAVA 91

RESULT 5
PSEB_CYAPA STANDARD; PRT; 38 AA.
ID PSEB_CYAPA
AC P48272;
DT 01-FEB-1996 (Rel. 33; Created)
DT 01-FEB-1996 (Rel. 33; Last sequence update)
DT 10-OCT-2003 (Rel. 42; Last annotation update)
DE Photosystem II protein Y.
OS Pseba OR YCF32.
OS Cyanophora paradoxa.
OC Cyanophora.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
ON NCBI_TaxID=2762;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Bryant D.A.;
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa";
RL Plant Mol. Biol. Rep. 13:327-332 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M., Farley J.Y., Schlucher W.M., Chung S., Newman-Spallart C., Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa: the genetic complexity of a primitive plastid.";
RL (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E., Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg (1997).
RN [3]
RP FUNCTION: Manganese-binding polypeptide with L-arginine metabolizing enzyme activity. Component of the core of photosystem II (by similarity).
CC -1- SUBCELLULAR LOCATION: Cyanelle thylakoid membrane.
CC -1- SIMILARITY: Belongs to the pseb family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
 CC or send an email to license@isb-sib.ch).

```

DR EMBL; U30821; AA81314.1; -.
DR PIR; T06971; T06971.
DR HMAP; ME 00717; -.
DR PHOTOSYNTHESIS; Photosystem II; Transmembrane; Thylakoid; Cyanelle.
FT DOMAIN 1 4 LUMENAL (POTENTIAL).
FT TRANSMEM 5 23 POTENTIAL.
FT DOMAIN 24 38 STROMAL (POTENTIAL).
SQ SEQUENCE 38 AA; 4240 MW; D11996A62BA74DFC CRC64;

Query Match 7.7%; Score 49; DB 1; Length 38;
Best local similarity 34.4%; Pred. No. 57;
Matches 11; Conservative 7; Mismatches 10; Indels 4; Gaps 1;

Oy 9 SLRVAKVAYPLGLCVGVPIY---VAVIKWHR 36
Db 2 SMRLVVLLPLGIALGWAVYNGIKLAEQWR 33

RESULT 6
GAST HUMAN
ID _GAST_HUMAN STANDARD; PRT; 101 AA.
AC P01350; P78463; P78464;
DT 21-JUL-1986 (Rel. 01, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Gastrin precursor.
GN GAS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87219893; PubMed=3034736;
RA Kariya Y., Kato K., Hayashizaki Y., Himeno S., Tarui S., Matsubara K.;
RT "Expression of human gastrin gene in normal and gastrinoma tissues.";
EL Gene 50:345-352(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=84272693; PubMed=6087340;
RA Ito R., Sato K., Helmer T., Jay G., Agarwal K.L.;
RT "Structural analysis of the gene encoding human gastrin: the large
RT intron contains an Alu sequence.";
EL Proc. Natl. Acad. Sci. U.S.A. 81:4662-4666(1984).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=84169471; PubMed=6324077;
RA Kato K., Hayashizaki Y., Takahashi Y., Himeno S., Matsubara K.;
RT "Molecular cloning of the human gastrin gene.";
EL Nucleic Acids Res. 11:8197-8203(1983).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=83221503; PubMed=6574456;
RA Boel E., Vuust J., Norris F., Norris K., Wind A., Rehfeld J.F.,
RA Marcker K.A.;
RT "Molecular cloning of human gastrin cDNA: evidence for evolution of
RT gastrin by gene duplication.";
EL Proc. Natl. Acad. Sci. U.S.A. 80:2866-2869(1983).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=84144842; PubMed=6322186;
RA Wiborg O., Berglund L., Boel E., Norris F., Norris K., Rehfeld J.F.,
RA Marcker K.A., Vuust J.;
RT "Structure of a human gastrin gene.";
EL Proc. Natl. Acad. Sci. U.S.A. 81:1067-1069(1984).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=84159488; PubMed=6689486;
RA Kato K., Himeno S., Takahashi Y., Makabayashi T., Tarui S.,
RA Matsubara K.;

```

```

RT "Molecular cloning of human gastrin precursor cDNA.";
RL Gene 26:53-57(1983).
RN [7]
RP SEQUENCE OF 22-101, AND CHARACTERIZATION OF GASTRIN 71.
RC TISSUE=Antral mucosa;
RX MEDLINE=94333379; PubMed=8055952;
RA Rehfeld J.F., Johnsen A.H.;
RT "Identification of gastrin component I as gastrin-71. The largest
RT possible bioactive progastrin product.";
RL Eur. J. Biochem. 223:765-773(1994).
RN [8]
RP SEQUENCE OF 76-92.
RX MEDLINE=67021327; PubMed=5921183;
RA Bentley P.H., Kenner G.W., Sheppard R.C.;
RT "Structure of human gastrins I and II.";
RL Nature 209:583-585(1966).
RN [9]
RP SEQUENCE OF 59-68.
RX MEDLINE=89273602; PubMed=2730647;
RA Higashimoto Y., Himeno S., Shinomura Y., Nagao K., Tamura T.,
RA Tarui S.;
RT "Purification and structural determination of urinary NH2-terminal
RT big gastrin fragments.";
RL Biochem. Biophys. Res. Commun. 160:1364-1370(1989).
RN [10]
RP SEQUENCE OF 76-92.
RX MEDLINE=69298172; PubMed=5822140;
RA Gregory R.A., Tracy H.J., Agarwal K.L., Grossman M.I.;
RT "Amino acid constitution of two gastrins isolated from
RT Zollinger-Ellison tumour tissue.";
RL Gut 10:603-608(1969).
CC -!- FUNCTION: Gastrin stimulates the stomach mucosa to produce and
CC secrete hydrochloric acid and the pancreas to secrete its
CC digestive enzymes. It also stimulates smooth muscle contraction
CC and increases blood circulation and water secretion in the stomach
CC and intestine.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- PTM: TWO DIFFERENT PROCESSING PATHWAYS PROBABLY EXIST IN ANTRAL
CC G-CELLS. IN THE DOMINANT PATHWAY PROGASTRIN IS CLEAVED AT THREE
CC SITES RESULTING IN TWO MAJOR BIOACTIVE GASTRINS, GASTRIN-34 AND
CC GASTRIN-17. IN THE PUTATIVE ALTERNATIVE PATHWAY, PROGASTRIN MAY
CC BE PROCESSED ONLY AT THE MOST C-TERMINAL DIABASIC SITE RESULTING
CC IN THE SYNTHESIS OF GASTRIN-71.
CC -!- SIMILARITY: Belongs to the gastrin/cholecystokinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

EMBL; X00183; CAA25005.1; -.
EMBL; X00183; CAA25006.1; -.
EMBL; X00183; CAA25007.1; -.
EMBL; V00511; CAA23769.1; -.
EMBL; M15958; AAA52520.1; -.
EMBL; K01254; AAB59533.1; -.
PIR; A91937; GMHUB.
DR Genew; HGNC:4164; GAS.
DR MIM; 137250; -.
DR GO; GO:0005180; F:peptide hormone; TAS.
DR InterPro; IPR001651; Gastrin.
DR Pfam; PF00918; Gastrin; 1.
DR SMART; SM00029; GASTRIN; 1.
DR PROSITE; PS00259; GASTRIN; 1.
KW Hormone; Cleavage on pair of basic residues; Amidation; Sulfation;
KW Signal; Phosphorylation; Pyrrolidone carboxylic acid.
FT SIGNAL 1 21
FT PEPTIDE 22 92
FT PEPTIDE 40 92
FT PEPTIDE 59 92
FT PEPTIDE 71 (COMPONENT I).
FT PEPTIDE 72 (POTENTIAL).
FT BIG GASTRIN (GASTRIN 34) (COMPONENT II).
```

```
FT PEPTIDE 76 92 GASTRIN (GASTRIN 17) (COMPONENT III).
FT MOD_RES 59 59 PYROGLUTAMIC ACID.
FT MOD_RES 76 59 PYROGLUTAMIC ACID.
FT MOD_RES 76 76 SULFATION (33%).
FT MOD_RES 87 87 AMIDATION (G-93 PROVIDE AMIDE GROUP).
FT MOD_RES 92 92 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 96 96 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 101 AA; 11394 MW; A03C847FCF7216C CRC64;

Query Match 7.7%; Score 49; DB 1; Length 101;
Best Local Similarity 35.3%; Pred. No. 1.7e+02;
Matches 18; Conservative 3; Mismatches 12; Indels 18; Gaps 3;

Oy 21 LCVGVFIVAYIKMHRATATQAFSITRAAPGAW---GQQAHSPLGTAD 68
Db 4 LCVVVLIF-----ALALAASE-----ASMKRSQDPALGTGAN 39

RESULT 7
VE4_BPv4 STANDARD; PRT; 107 AA.
AC P08347;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1988 (Rel. 08, Last annotation update)
DE Probable B4 protein.
OS Bovine Papillomavirus type 4.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
CX NCBI_TaxID=10562;
RN 11;
RP SEQUENCE FROM N.A.
RA MEDLINE=87282264; PubMed=3039043;
RA Patel K.R., Smith K.T., Campo M.S.;
RT "The nucleotide sequence and genome organization of bovine
RT Papillomavirus type 4";
RT J. Gen. Virol. 68:2117-2128 (1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; X05817; NOT ANNOTATED_CDS.
DR PIR; G27129; W4WLB4.
KW Early protein.
SQ SEQUENCE 107 AA; 11846 MW; 1CC2A84EF30B4C2 CRC64;

Query Match 7.6%; Score 48.5; DB 1; Length 107;
Best Local Similarity 28.1%; Pred. No. 2e+02;
Matches 16; Conservative 10; Mismatches 20; Indels 11; Gaps 3;

Oy 13 AKVAVPLGLCGVFIYVAY---IKMHRATATQAFSITRAAPGAWGQQAHSPLGT 65
Db 48 AEGAVVPAIATVITITVEGALQWTRHATE-----SETPGRK-TGHPRTGT 97

RESULT 8
GDIR_CAVPO STANDARD; PRT; 111 AA.
AC P80237;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Rho GDI-disassociation inhibitor 1 (Rho GDI 1) (Rho-GDI alpha)
DE (Fragments).
GN ARHGDI1.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Heteriognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
```

```
RN [1]
RP STRAIN=Hartley; TISSUE=Macrophage;
RC MEDLINE=94039069; PubMed=8223583;
RX Pick E., Gorzalczyk Y., Engel S.;
RA "Role of the rac1 p21-GDP-dissociation inhibitor for rho heterodimer
RT in the activation of the superoxide-forming NADPH oxidase of
RT macrophages";
RT Eur. J. Biochem. 217:441-455 (1993).
CC -1- FUNCTION: Regulates the GDP/GTP exchange reaction of the Rho
CC proteins by inhibiting the dissociation of GDP from them, and the
CC subsequent binding of GTP to them.
CC -1- FUNCTION: Rac1 p21/Rho GDI heterodimer is the active component of
CC the cytosolic factor sigma 1, which is involved in stimulation of
CC the NADPH oxidase activity in macrophages.
CC -1- SUBUNIT: FORMS A HETERODIMER WITH P21/RAC-1.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- PTM: The N-terminus is blocked.
CC -1- SIMILARITY: Belongs to the Rho GDI family.
DR HSSP; P19803; 1GDF.
DR InterPro; IPR007110; I9-1like.
DR InterPro; IPR000406; Rho_GDI.
DR PRINTS; PR00492; RHODI.
KW GTPase activation.
FT NON_TER 1
FT NON_CONS 40 41
FT NON_CONS 46 47
FT NON_CONS 86 87
FT UNSURE 98 98
FT UNSURE 100 100
FT NON_TER 111
SQ SEQUENCE 111 AA; 12496 MW; 0FCE35BEA40FE951 CRC64;

Query Match 7.6%; Score 48.5; DB 1; Length 111;
Best Local Similarity 26.8%; Pred. No. 2.1e+02;
Matches 15; Conservative 5; Mismatches 21; Indels 15; Gaps 2;

Oy 70 HEVTFGIMFD---AGSTGTRVHVFOFTRPREPTITHEFKVKGGLSYADD 120
Db 50 HTRYKGVKIDKTDYVGVGYPRAEYEFILTPWEAP-----KQMLAFSTD 95

RESULT 9
PSBE_CHLVU STANDARD; PRT; 83 AA.
ID PSBE_CHLVU
AC P56309;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome b559 alpha subunit (PSII reaction center subunit V).
GN PSBE.
OS Chlorococcidiaceae.
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorococcales;
OC Chlorococcaceae; Chlorococcia.
OX NCBI_TaxID=3077;
RN 11;
RP SEQUENCE FROM N.A.
RC STRAIN=IAM C-27 / Tamiya;
RX MEDLINE=97303241; PubMed=9159184;
RA Wakaue T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,
RA Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,
RA Inamura A., Yoshinaga K., Sugita M.;
RT "Complete nucleotide sequence of the chloroplast genome from the
RT green alga Chlorococcidiaceae: the existence of genes possibly
RT involved in chloroplast division";
RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972 (1997).
CC -1- FUNCTION: This b-type cytochrome is tightly associated with the
CC reaction center of photosystem II and possibly is part of the
CC water-oxidation complex.
CC -1- SUBUNIT: Heterodimer of an alpha subunit and a beta subunit.
CC -1- SIMILARITY: Belongs to the psb2 / psbP family.
CC -----
```



```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB001684; BAA57904.1; -
CC PIR; T07257; T07257.
CC DR HAMAP; MF 00642; -; 1.
CC DR InterPro; IPR006216; Cyt_b559.
CC DR InterPro; IPR006217; Cyt_b559_alpha.
CC DR Pfam; PF00283; cytochr_b559; 1.
CC DR Pfam; PF00284; cytochr_b559a; 1.
CC DR ProDom; PD004478; Cyt_b559_alpha; 1.
CC DR TIGRfam; TIGR01332; Cyt_b559_alpha; 1.
CC DR PROSITE; PS00537; CYTOCHROME B559; 1.
CC DR Chloroplast; Photosystem II; Heme; Electron transport; Transmembrane.
CC DONAIN 1 38 STROMAL (POTENTIAL).
CC FT TRANSMEM 19 39 POTENTIAL.
CC FT DOMAIN 40 83 LUMENAL (POTENTIAL).
CC FT METAL 23 23 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC SQ SEQUENCE 83 AA; 9476 MW; 78FDEP48EFF7D911 CRC64;
CC -----
Query Match 7.58; Score 48; DB 1; Length 83;
Best Local Similarity 32.14; Pred. No. 1.7e+02;
Matches 17; Conservative 7; Mismatches 17; Indels 12; Gaps 3;
Qy 70 HEVFGYGMFDAG-----STGRVHVF-----QTPRPRTPLTHETPKAVK 111
Db 23 HSIPTSLFTAGLWLVSTGLAYDFGSPRNEYFTEDRQETPLIT-DRFNAL 74
CC -----
RESULT 10
VPP_VPPRD STANDARD; PRT; 83 AA.
AC P27391.
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Protein P (GPP).
GN P.
OS Bacteriophage PRD1.
OC Viruses; dsDNA viruses, no RNA stage; Tectiviridae; Tectivirus.
OC NCBI_TaxID=10658;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-9.
RX MEDLINE=91306449; PubMed=1853567;
RA Bamford J.K.H., Haeninen A.-J., Pakula T.M., Ojala P.M.,
RA Kalkinen N., Frilander M., Bamford D.H.;
RT "Genome organization of membrane-containing bacteriophage PRD1.";
PL Virology 183:658-676 (1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M69077; AAA32471.1; -
CC PIR; H36776; WMBPWB.
CC DR INIT MET 0
CC SQ SEQUENCE 83 AA; 9055 MW; C61CEAD11098378C CRC64;
CC -----
Query Match 7.43; Score 47.5; DB 1; Length 83;
Best Local Similarity 34.34; Pred. No. 2e+02;
Matches 12; Conservative 5; Mismatches 15; Indels 3; Gaps 1;
Qy 35 HRATATQAFSITRAAPGARWGQQAHSPLGTAADG 69
Db 35 HRATATQAFSITRAAPGARWGQQAHSPLGTAADG 69
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF063097; AAO03283.1; -
CC DR InterPro; IPR007048; GPW_gp25.
CC DR Pfam; PF04965; GPW_gp25; 1.
CC SQ SEQUENCE 115 AA; 12637 MW; F2684FB3DAA7DE57 CRC64;
CC -----
Query Match 7.43; Score 47.5; DB 1; Length 115;
Best Local Similarity 33.38; Pred. No. 2.8e+02;
Matches 18; Conservative 10; Mismatches 15; Indels 11; Gaps 4;
Qy 1 MKKISNHSGLRVAKVAYP---LGLCVGVFTYVAYIKWH-RATATQAFSITRA 49
Db 39 MRR--DYGSLIASMIDPQTPALELQIKVACVMAVLKWEPRVILS-----SVTTA 86
CC -----
RESULT 11
VPP_VPP2 STANDARD; PRT; 115 AA.
AC P51768;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Baseplate assembly protein W (GPW).
GN W.
OS Bacteriophage P2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC P2-like viruses.
OC NCBI_TaxID=10679;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96036485; PubMed=7483254;
RA Hagsaard-Ljungquist E., Jacobsen E., Rishovd S., Six E.W., Nilssen O.,
RA Sunshine M.G., Lindqvist B.H., Kim K.-J., Barreiro V., Koonin E.V.,
RA Calender R.;
RT "Bacteriophage P2: genes involved in baseplate assembly.";
RL Virology 213:109-121 (1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF063097; AAO03283.1; -
CC DR InterPro; IPR007048; GPW_gp25.
CC DR Pfam; PF04965; GPW_gp25; 1.
CC SQ SEQUENCE 115 AA; 12637 MW; F2684FB3DAA7DE57 CRC64;
CC -----
Query Match 7.43; Score 47.5; DB 1; Length 115;
Best Local Similarity 33.38; Pred. No. 2.8e+02;
Matches 18; Conservative 10; Mismatches 15; Indels 11; Gaps 4;
Qy 1 MKKISNHSGLRVAKVAYP---LGLCVGVFTYVAYIKWH-RATATQAFSITRA 49
Db 39 MRR--DYGSLIASMIDPQTPALELQIKVACVMAVLKWEPRVILS-----SVTTA 86
CC -----
RESULT 12
YC9C_SCHPO STANDARD; PRT; 115 AA.
AC Q09886;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C584.12 in chromosome III.
GN SPCC584.12.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Haidago J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
```

```

CC RA Oliver K.O'Neil S., Pearson D., Quail M.A., Rabinovitch E.,
CC RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
CC RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
CC RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
CC RA Woodward J., Volckaert G., Aert R., Robben J., Grymoult B.,
CC RA Weltjens I., Vanterleers E., Rieger M., Schaefer M., Weiller-Auer S.,
CC RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moeserl D., Hilbert H.,
CC RA Borzym K., Langner I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
CC RA Eger P., Zimmermann W., Medler H., Mambrt R., Purnelle B.,
CC RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motlier S.,
CC RA Lucban F., Aves S.J., Xiang Z., Hunt C., Moore K., Huret S.M.,
CC RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
CC RA Daga R.R., Cruzado I., Jimenez J., Sanchez M., del Rey F., Bento J.,
CC RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
CC RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
CC RA Shpakovskii G.V., Uesery D., Barrell B.G., Nurse P.:
CC RT "The genome sequence of Schizosaccharomyces pombe."
CC RL Nature 415:871-880(2002).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AL032824; CAB37425.1; -
CC DR PIR: T39132; S62526
CC DR GeneDB: Spombe; SPC0584.12; -
CC DR Hypothetical protein.
CC KQ SEQUENCE 115 AA; 13701 MW; 289D50E9AE5CA13E CRC64;
CC SQ
CC
CC Query Match 7.4%; Score 47.5; DB 1; Length 115;
CC Best Local Similarity 63.2%; Pred. No. 2.8e+02;
CC Matches 12; Conservative 0; Mismatches 6; Indels 1; Gaps 1
CC
CC QY 93 TRPPREPTLHETPKAVK 111
CC Db 7 TFPRRRRPTL-HEITENHYK 24
CC
CC
CC RESULT 13
CC YAHH_ECOLI
CC AC P75690;
CC ID YAHH_ECOLI STANDARD; PRT; 106 AA.
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Hypothetical protein yahh.
CC GN YAHH OR B0322.
CC OS Escherichia coli.
CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC OC Enterobacteriaceae; Escherichia.
CC OX NCBI_Taxid=562;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=K12 / MG1655;
CC RX MEDLINE=97426617; PubMed=9278503;
CC RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
CC RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
CC RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
CC RA Mau B., Shao Y.:
CC RT "The complete genome sequence of Escherichia coli K-12."
CC RL Science 277:1453-1474(1997).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).

```

```

CC      EMBL; AB000139; AAC73425.1; -.
DR      PIR; B64759; B64759.
DR      Ecogene; EG13592; yanh.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 106 AA; 10634 MW; 2260B0C42954AF71 CRC64;

Query Match          7.3%; Score 47; DB 1; Length 106;
Best Local Similarity 24.7%; Pred. No. 2.9e+02;
Matches 20; Conservative 13; Mismatches 32; Indels 16; Gaps 4;

QY      37 ATATQAFFST---TRAPGARWQQAHPVGTADGHEVYVYGMFDAGSTGRVHYVFOF 92
DB      18 SVAAQAFITLIGCCVNMALSGIRMAQES---VGLISVAHQAF-----ATTAGCGVDALSG 68
QY      93 TRPRETP---TLTHETPKAV 110
DB      69 LRVARBSVGLISVAHQAFVTI 89

RESULT 14
NT3M_CYPCA          STANDARD; PRT; 116 AA.
AC      P24974;
DT      01-MAR-1992 (Rel. 21, Created)
DT      01-MAR-1992 (Rel. 21, Last sequence update)
DE      16-OCT-2001 (Rel. 40, Last annotation update)
DE      NADH-ubiquinone oxidoreductase chain 3 (EC 1.6.5.3).
DS      NTND3 OR ND3.
OS      Cyprinus carpio (Common carp).
OC      Mitochondrion.
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC      Cyprinidae; Cyprinus.
OX      NCBI_TaxID=7962;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=94223691; PubMed=8169959;
RA      Chang Y.S., Huang F.L., Lo T.B.;
RT      "The complete nucleotide sequence and gene organization of carp
RL      (Cyprinus carpio) mitochondrial genome.";
RJ      J. Mol. Evol. 38:138-155(1994).
CC      -1- CATABOLIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC      -1- SIMILARITY: Belongs to the complex I subunit 3 family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on items
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X61010; CA43334.1; -.
DR      PIR; S36004; S36004.
DR      InterPro; IPR000440; Oxidored q4.
DR      Pfam; PF00507; oxidored q4; 1.
KW      Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ      SEQUENCE 116 AA; 13071 MW; DFE5E9F4D1DEFAER CRC64;

Query Match          7.3%; Score 47; DB 1; Length 116;
Best Local Similarity 63.6%; Pred. No. 3.2e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      55 WGGQAHPPLGT 65
DB      78 WGDOLHNPPTG 88

RESULT 15
ID      Y109_EMENTI          STANDARD; PRT; 97 AA.
AC      P05678;

```

Search completed: June 8, 2004, 10:00:48
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 09:57:40 ; Search time 39 Seconds
(without alignments)
970.825 Million cell updates/sec

Title: US-09-905-589a-2_copy_1_120
Perfect score: 640
Sequence: 1 MKRISNHSGLRVAKVAVPLD.....TLTHTFKAVKQLSAVDD 120

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 231205

Minimum DB seg length: 0
Maximum DB seg length: 120

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75.5	11.8	119	16	Q7UNK7 rhodopsin
2	69.5	10.9	64	12	Q8Q1L2 rhodopsin
3	60.5	9.5	105	2	Q83ZP7 escherichia
4	60	9.4	106	3	Q86ZEL phoma sp. c
5	60	9.4	114	16	Q91LQ3 streptomyces
6	59	9.2	102	8	Q9TEV4 opsinus bet
7	58.5	9.1	102	8	Q9T3H8 opsinus bet
8	58.5	9.1	102	8	Q9TEV3 opsinus par
9	58.5	9.1	102	8	Q9TEV8 porichnys
10	58	9.1	99	5	Q9STU9 opsinus tau
11	57.5	9.0	102	8	Q9TEV1 opsinus tau
12	57.5	9.0	102	8	Q9TEV2 opsinus tau
13	57.5	9.0	102	8	Q9B272 rhodopsin
14	57.5	9.0	105	8	Q9B272 rhodopsin
15	57.5	9.0	120	8	Q9T7C8 rhodopsin
16	57	8.9	70	16	Q8XWKL talastonia s

17	57	8.9	108	16	Q8NPP8 corynebacte
18	56.5	8.8	91	16	Q8ZUV9 nitrosomna
19	56.5	8.8	101	8	Q20821 lerista bou
20	56.5	8.8	102	8	Q20818 lerista bou
21	56.5	8.8	104	8	Q20820 lerista bou
22	56.5	8.8	105	8	Q9T7B6 trinomys ih
23	56.5	8.8	111	8	Q9T7B6 trinomys ih
24	56.5	8.8	118	8	Q9T7B7 trinomys ih
25	56	8.8	105	8	Q9B268 pachyactyl
26	56	8.8	115	17	Q9Y299 aeropyrum p
27	55.5	8.7	100	8	Q47715 cephalopnus
28	55.5	8.7	111	8	Q9MEJ5 pronolagus
29	55	8.6	96	8	Q9B430 hemidactylu
30	54.5	8.5	73	16	Q9RD85 streptomyce
31	54.5	8.5	85	16	Q7U6D9 streptomyce
32	54.5	8.5	105	8	Q8WD29 xantusia he
33	54.5	8.5	108	8	Q9T7D1 trinomys se
34	54.5	8.5	111	8	Q9MEJ9 pronolagus
35	54.5	8.5	119	8	Q9MEK0 propithecus
36	54.5	8.5	119	8	Q9MEK8 lepitemur s
37	54	8.4	101	10	Q7XUB5 oryza sativ
38	54	8.4	119	16	Q9L220 streptomyce
39	53.5	8.4	67	16	Q87R21 vibrio para
40	53.5	8.4	96	8	Q86050 tupia glis
41	53.5	8.4	98	8	Q85Q60 hippidion s
42	53.5	8.4	104	8	Q9T7D5 trinomys se
43	53.5	8.4	111	8	Q9MEJ8 pronolagus
44	53.5	8.4	113	16	Q8ZBW0 yersinia pe
45	53.5	8.4	116	8	Q9T7D6 trinomys se

ALIGNMENTS

RESULT 1

Q7UNK7 PRELIMINARY; PRT; 119 AA.

AC Q7UNK7;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN R57531
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxId=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=; PubMed=12835416;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heilmann K., Rabus R.,
RA Schleiner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
strain 1."
RT Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
RL EMBL: BX294146; CAD75412.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 119 AA; 13173 MW; 9BAEA8D8C6FFB8 CRC64;

Query Match 11.8%; Score 75.5; DB 16; Length 119;
Best Local Similarity 30.3%; Pred. No. 1;
Matches 36; Conservative 11; Mismatches 49; Indels 23; Gaps 8;

QY 6 NFGSLRVAKVAVPLDLCGVETIYVAYIKMRATATQAF-ESITRAPG---ARMGQQAHS 61
DB 4 NSGDASKARRACPOHSGTGLIRVKSTIS---GLATIGDPCTIVNAKGLVRSQVQVPHR 60
QY 62 PLG-TAADQHE-VFQGIWEDA-----GSTGT---RVHVFQFTRPP-----RETPTLTH 104
DB 61 PLGSLRVAKVAVPLDLCGVETIYVAYIKMRATATQAF-ESITRAPG---ARMGQQAHS 61

```
RESULT 2
Q8QLI2 PRELIMINARY; PRT; 64 AA.
AC Q8QLI2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mamestra configurata nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=191492;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=90/2;
RX MEDLINE=97163493; PubMed=9010313;
RA Li S., Erlandson M., Moody D., Gillett C.;
RT "A physical map of the Mamestra configurata nucleopolyhedrovirus
RT genome and sequence analysis of the polyhedrin gene.";
RL J. Gen. Virol. 78:265-271(1997).
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN=90/2;
RX MEDLINE=21884635; PubMed=11886270;
RA Li Q., Donly C., Li L., Willis L.G., Theilmann D.A., Erlandson M.;
RT "Sequence and Organization of the Mamestra configurata
RT Nucleopolyhedrovirus Genome.";
RL Virology 294:106-121(2002).
RN [3];
RP SEQUENCE FROM N.A.
RC STRAIN=90/2;
RA Li Q., Donly C., Li L., Willis L.G., Theilmann D.A., Erlandson M.A.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U59461; AAM09160.1; -.
KW Hypothetical protein.
SQ SEQUENCE 64 AA; 7844 MW; 472234944556505A CRC64;

Query Match 10.9%; Score 69.5; DB 12; Length 64;
Best Local Similarity 33.3%; Pred. No. 2.3;
Matches 14; Conservative 10; Mismatches 11; Indels 7; Gaps 2;

QY 2 RKISNHSRLRVAKVAPLGLCV-----GVFIYVAYIKWHRA 37
Db 14 RVTINHKVHKVKNRMP-FICIIHPQCSWYIVQYIRWHS 54

RESULT 3
Q83ZP7 PRELIMINARY; PRT; 105 AA.
AC Q83ZP7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative small multidrug resistance protein.
GN SMR.
OS Escherichia coli.
OG Plasmid pAK33.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1];
RP SEQUENCE FROM N.A.
RA Vourli S., Tzouveleakis L.S., Tzelepi E., Lebesse E., Legakis N.J.,
RA Miragou V.;
RT "Integron In11, a novel class I integron from Escherichia coli, which
RT is bounded by IS26 insertion elements.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY260546; AAP22975.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR InterPro; IPR000390; Smr.
DR Pfam; PF00893; Multi_Drug_Res; 1.

KW Plasmid.
SQ SEQUENCE 105 AA; 11210 MW; 30B57F9D4C73884 CRC64;

Query Match 9.5%; Score 60.5; DB 2; Length 105;
Best Local Similarity 25.0%; Pred. No. 45;
Matches 18; Conservative 12; Mismatches 29; Indels 13; Gaps 3;

QY 14 KVAYPLGLCVGVFIYVAYIKWHRAATATQAFSTRAAPGARWQQQAHSPLGTADGHEVF 73
Db 13 EIGWPGVLKMA---QVPETRWSGVGIAPVAVSGF---LLMLAQRHPIGTA----- 59
QY 74 YGIMFDAGSTGT 85
Db 60 YAVWTGIGAGT 71

RESULT 4
Q86ZE1 PRELIMINARY; PRT; 106 AA.
AC Q86ZE1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE C-methyl transferase (fragment).
OS Phoma sp. C2932.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Phoma.
OX NCBI_TaxID=86977;
RN [1];
RP SEQUENCE FROM N.A.
RX MEDLINE=21150635; PubMed=11251290;
RA Nicholson T.P., Rudd B.A.M., Dawson M., Lazarus C.M., Simpson T.J.,
RA Cox R.J.;
RT "Design and utility of oligonucleotide gene probes for fungal
RT polyketide synthases.";
RL Chem. Biol. 8:157-178(2001).
RN [2];
RP SEQUENCE FROM N.A.
RA Glod F., Cox R.J., Simpson T.J.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY217787; AAC62424.1; -.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR001601; Methyltransf.
DR InterPro; IPR000051; SAM_bind.
KW Transferase.
FT NON TER 1
FT NON TER 106
SQ SEQUENCE 106 AA; 11440 MW; E9A1D80A0E2EE53E CRC64;

Query Match 9.4%; Score 60; DB 3; Length 106;
Best Local Similarity 36.8%; Pred. No. 52;
Matches 15; Conservative 4; Mismatches 22; Indels 0; Gaps 0;

QY 80 AGSTGTRVHVQFTRPRPTTTLTHTFKAVKPGLSAYADD 120
Db 4 AGTGGTYHVLRLRNPDGSSKATQVHTDISPGFLAKAD 44

RESULT 5
Q9LIQ3 PRELIMINARY; PRT; 114 AA.
AC Q9LIQ3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein SCO6862.
GN SCO6862 OR SC7F9.14C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1];
RP SEQUENCE FROM N.A.
```

RC STRAIN-A3(2) / M145;
RX MEDLINE-21996410; PubMed=1200953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Godle A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajadream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wierozorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939129; CAB72367.1; -
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 114 AA; 12316 MW; 1BA5C6423312E16D CRC64;

Query Match 9.4%; Score 60; DB 16; Length 114;
Best Local Similarity 28.9%; Pred. No. 57;
Matches 22; Conservative 11; Mismatches 31; Indels 12; Gaps 3;

QY 33 KMHK---ATATGAFSITPAAPGAWGQQAHSPLGTADGHEVFQIMPDAGSTGRVAV 89
DB 22 QMRODTARTETITLQSAKP-----AGEGMDALDVIALVQSGSTASLVLTIVA 74

QY 90 PQFTRPPEPTLTHE 105
DB 75 WRTRP--STPVVTE 88

RESULT 6

Q9TEV4 PRELIMINARY; PRT; 102 AA.
AC Q9TEV4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome b (Fragment).
GN C1TB OR C1T-B.
OS Opanus beta.
OG Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Batrachoididae; Opanus.
OX NCBI_TaxID=95145;
RN [1]
RP SEQUENCE FROM N.A.
RA Freshwater D.W., Kyhn-Hansen C., Sarver S.K., Walsh P.J.;
RT "Phylogeny of Opanus (Batrachoididae) inferred from multiple
RT mitochondrial DNA sequences.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RISKS PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
DR EMBL; AF165346; AAD48462.1; -
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR005797; Cytb_b6_N.
DR Pfam; PF00033; cytochrome_b_N; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
KM Electron transport; Heme; Respiratory chain; Transmembrane; Transport;
KM Mitochondrion.

FT NON TER 1 1
FT NON TER 102 102
SQ SEQUENCE 102 AA; 11494 MW; B0B5D97908F301AA CRC64;

Query Match 9.2%; Score 59; DB 8; Length 102;
Best Local Similarity 28.8%; Pred. No. 64;
Matches 23; Conservative 11; Mismatches 18; Indels 28; Gaps 5;

QY 19 LGLCV-----GVFIYAYIKMRATATQAFSITPAAPGAWG---QQAHSPLGTADG 69
DB 5 LGMCLLVQMTTGVFIAMHYI-----PDATMAFPISLAHSRDVNYGWLIRNTH-----ANG 54

QY 70 HEVRY-----GIMFDA 80
DB 55 ASVFEMCYLHIGRIYDS 74

RESULT 7

Q9T3H8 PRELIMINARY; PRT; 102 AA.
AC Q9T3H8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome b (Fragment).
GN C1TB OR C1T-B.
OS Opanus beta.
OG Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Batrachoididae; Opanus.
OX NCBI_TaxID=95145;
RN [1]
RP SEQUENCE FROM N.A.
RA Freshwater D.W., Kyhn-Hansen C., Sarver S.K., Walsh P.J.;
RT "Phylogeny of Opanus (Batrachoididae) inferred from multiple
RT mitochondrial DNA sequences.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RISKS PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
DR EMBL; AF165345; AAD48461.1; -
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR005797; Cytb_b6_N.
DR Pfam; PF00033; cytochrome_b_N; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
KM Electron transport; Heme; Respiratory chain; Transmembrane; Transport;
KM Mitochondrion.

FT NON TER 1 1
FT NON TER 102 102
SQ SEQUENCE 102 AA; 11450 MW; B0B5D97908F651FF CRC64;

Query Match 9.1%; Score 58.5; DB 8; Length 102;
Best Local Similarity 30.4%; Pred. No. 73;
Matches 21; Conservative 10; Mismatches 19; Indels 19; Gaps 4;

QY 19 LGLCV-----GVFIYAYIKMRATATQAFSITPAAPGAWG---QQAHSPLGTADG 69
DB 5 LGMCLLVQMTTGVFIAMHYI-----PDATMAFPISLAHSRDVNYGWLIRNTH-----ANG 54
QY 70 HEVRYGIMF 78

```

OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Batrachoididae; Forichthys.
NCBI_TaxID=45384;
[1]
RN SEQUENCE FROM N.A.
RA Freshwater D.W., Kyhn-Hansen C., Sarver S.K., Walsh P.J.;
RT "Phylogeny of Opeanus (Batrachoididae) inferred from multiple
RL mitochondrial DNA sequences.";
RC Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -I- COPACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -I- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC EMBL; AF165352; AAD48468.1; -.
CC GO; GO:0014021; C:integral to membrane; IEA.
CC GO; GO:0005746; C:mitochondrial electron transport chain; IEA.
CC GO; GO:0005739; C:mitochondrion; IEA.
CC GO; GO:0016491; F:oxygen reductase activity; IEA.
CC GO; GO:0006118; P:electron transport; IEA.
CC GO; GO:0006810; P:transport; IEA.
CC InterPro: IPR005797; Cytb_b6_N.
CC Pfam: PF000033; cytochrome_b_N; 1.
CC PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
KW Electron transport; Heme; Respiratory chain; Transmembrane; Transport;
KW Mitochondrion.
FT NON_TER 1
FT TER 102
SQ SEQUENCE 102 AA; 11357 MW; 4E9DA2A8ADCE1B84 CRC64;

Query Match 9.1%; Score 58.5; DB 8; Length 102;
Best Local Similarity 30.4%; Pred. NO. 73;
Matches 21; Conservative 9; Mismatches 20; Indels 19; Gaps 4;

QY 19 LGLCV-----GVFIYVAVIKWHRATATQAFFSITRAAFGARWG---QQAHSLPLGTAADG 69
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5 LGLCLVTQMIGTVFIAMHYI----PDALAFSSVAHLTRDVNYGWLRLNLH-----ANG 54

QY 70 HEVFYGIMF 78
:: :: ::
Db 55 ASIFPICMY 63

RESULT 10
Q9STD9 PRELIMINARY; PRT; 99 AA.
AC Q9STD9;
DT 01-DRC-2001 (TrEMBLrel. 19, Created)
DT 01-DRC-2001 (TrEMBLrel. 19, Last sequence update)
DT DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE LD29295P.
GN CG14211.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Berkely;
RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise B., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunco J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY059453; AAL13359.1; -.

```


DR FlyBase; FBgn031044; CG14211.
SQ SEQUENCE 99 AA; 11674 MW; B8F1CA2E34484680 CRC64;

Query Match 9.1%; Score 58; DB 5; Length 99;
Best Local Similarity 23.5%; Pred. No. 81;
Matches 16; Conservative 15; Mismatches 25; Indels 12; Gaps 2;

QY 2 RKSNHSLRVAKYAFGLCVGVIYVYIKMR-----ATATQAFSTTRAPGARWG 56
DB 16 RKTQHPSMTISK-----YALYIHITLYNNIVGQDTISVMSMERSPPASWG 68
QY 57 QOASPLG 64
DB 69 LRAPSDIG 76

RESULT 11

Q9TEVO PRELIMINARY; PRT; 95 AA.

AC Q9TEV2; 01-MAY-2000 (TRENBLREL.13, Created)
DT 01-MAY-2000 (TRENBLREL.13, Last sequence update)
DT 01-JUN-2003 (TRENBLREL.24, Last annotation update)
DE Cytochrome b (Fragment).
GN CYTB OR CYT-B.
OS Oposanus tau (Oyster toadfish).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphi; Paracanthopterygii; Batrachoidae; Oposanus.
OX NCBI_TaxID=8068;
RN [1]
RP SEQUENCE FROM N.A.
RA Freshwater D.W., Kyhn-Hansen C., Sarver S.K., Walsh P.J.;
RT "Phylogeny of Oposanus (Batrachoididae) inferred from multiple
mitochondrial DNA sequences."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
DR EMBL; AF165350; AAD48466.1; -
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR005797; Cytb b6 N.
DR Pfam; PF00033; Cytochrome_b_N.1
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
KW Electron transport; Heme; Respiratory chain; Transmembrane; Transport;
KM Mitochondrion.
FT NON TER 1
FT NON TER 95
SQ SEQUENCE 95 AA; 10763 MW; 8A53F225395E51FB CRC64;

Query Match 9.0%; Score 57.5; DB 8; Length 95;
Best Local Similarity 29.0%; Pred. No. 87;
Matches 20; Conservative 11; Mismatches 19; Indels 19; Gaps 4;

QY 19 LGLCV-----GVFIYVYIKMRATQAFSSITRAPGARWG---QOASPLGTADG 69
DB 1 LGMCLLVQMATGVFIAMRYI-----PDATMAFPSLAHLSRDVNYGLNLNHH-----ANG 50
QY 70 HEVFIYGVF 78
DB 51 ASIFMCMY 59

RESULT 12

Q9TEV1 PRELIMINARY; PRT; 102 AA.

AC Q9TEV1; 01-MAY-2000 (TRENBLREL.13, Created)
DT 01-MAY-2000 (TRENBLREL.13, Last sequence update)
DT 01-JUN-2003 (TRENBLREL.24, Last annotation update)
DE Cytochrome b (Fragment).
GN CYTB OR CYT-B.
OS Oposanus tau (Oyster toadfish).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphi; Paracanthopterygii; Batrachoidae; Oposanus.
OX NCBI_TaxID=8068;
RN [1]
RP SEQUENCE FROM N.A.
RA Freshwater D.W., Kyhn-Hansen C., Sarver S.K., Walsh P.J.;
RT "Phylogeny of Oposanus (Batrachoididae) inferred from multiple
mitochondrial DNA sequences."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
DR EMBL; AF165349; AAD48465.1; -
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR005797; Cytb b6 N.
DR Pfam; PF00033; Cytochrome_b_N.1
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
KW Electron transport; Heme; Respiratory chain; Transmembrane; Transport;
KM Mitochondrion.
FT NON TER 1
FT NON TER 102
SQ SEQUENCE 102 AA; 11430 MW; 0A00EF34A0ED687B CRC64;

Query Match 9.0%; Score 57.5; DB 8; Length 102;
Best Local Similarity 29.0%; Pred. No. 95;
Matches 20; Conservative 11; Mismatches 19; Indels 19; Gaps 4;

QY 19 LGLCV-----GVFIYVYIKMRATQAFSSITRAPGARWG---QOASPLGTADG 69
DB 5 LGMCLLVQMATGVFIAMRYI-----PDATMAFPSLAHLSRDVNYGLNLNHH-----ANG 54

RESULT 13

Q9TEV2 PRELIMINARY; PRT; 102 AA.

AC Q9TEV2; 01-MAY-2000 (TRENBLREL.13, Created)
DT 01-MAY-2000 (TRENBLREL.13, Last sequence update)
DT 01-JUN-2003 (TRENBLREL.24, Last annotation update)
DE Cytochrome b (Fragment).
GN CYTB OR CYT-B.
OS Oposanus phobeton.
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acanthomorpha; Paracanthopterygii; Batrachoididae; Opsanus.
 CC NCBI_TaxID=101190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Freshwater D.W., Kynn-Hansen C., Sarver S.K., Walsh P.J.;
 RT "Phylogeny of Opsanus (Batrachoididae) inferred from multiple
 RL mitochondrial DNA sequences.";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
 CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
 CC BOUND TO THE PROTEIN (BY SIMILARITY).
 CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 DR EMBL: AF165348; AAD48464.1; --
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0005746; C:mitochondrial electron transport chain; IEA.
 DR GO: GO:0005739; C:mitochondrion; IEA.
 DR GO: GO:0016491; P:oxidoreductase activity; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.
 DR GO: GO:0006810; P:transport; IEA.
 DR InterPro: IPR005797; Cytb b6 N.
 DR Pfam: PF00033; cytochrome_b_N; 1.
 DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
 KW Electron transport; Heme; Respiratory chain; Transmembrane; Transport;
 KW Mitochondrion.
 FT NON_TER 1
 FT NON_TER 102 102
 FT SEQUENCE 102 AA; 11411 MW; 6BD5D97908EF784D CRC64;
 QY Query Match 9.0%; Score 57.5; DB 8; Length 102;
 Best Local Similarity 29.0%; Pred. No. 95;
 Matches 20; Conservative 11; Mismatches 19; Indels 19; Gaps 4;
 QY 19 LGLCV-----GVFIYVAYIKWHRATATQAFPSITRAAPGARWG---QQAHSPLGTADG 69
 Db 5 LGMCLLVQMTGCVFIAMHYI-----PDATMAPPSLAHLSRDVNYGMLLNH-----ANG 54
 QY 70 HEVFGIMF 78
 Db 55 ASIFFNCMY 63
 RESULT 14
 Q9B272 PRELIMINARY; PRT; 105 AA.
 AC Q9B272;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Cytochrome b (Fragment).
 GN CYTB.
 OS Rhoptropus Boultoni.
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Lepidosauria; Squamata; Scleroglossa; Gekkota; Gekkonidae; Rhoptropus.
 CC NCBI_TaxID=152588;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lamb T., Bauer A.M.;
 RT "Mitochondrial phylogeny of Namib day geckos (Rhoptropus) based on
 RL cytochrome b and 16S rRNA sequences.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
 CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
 CC BOUND TO THE PROTEIN (BY SIMILARITY).

CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 DR EMBL: AY026922; AAK13033.1; --
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0005746; C:mitochondrial electron transport chain; IEA.
 DR GO: GO:0005739; C:mitochondrion; IEA.
 DR GO: GO:0016491; P:oxidoreductase activity; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.
 DR GO: GO:0006810; P:transport; IEA.
 DR InterPro: IPR005797; Cytb b6 N.
 DR Pfam: PF00033; cytochrome_b_N; 1.
 DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
 KW Electron transport; Heme; Respiratory chain; Transmembrane; Transport;
 KW Mitochondrion.
 FT NON_TER 1
 FT NON_TER 105 105
 FT SEQUENCE 105 AA; 11701 MW; 16D0905FAL2AD39F CRC64;
 QY Query Match 9.0%; Score 57.5; DB 8; Length 105;
 Best Local Similarity 30.8%; Pred. No. 99;
 Matches 24; Conservative 9; Mismatches 18; Indels 27; Gaps 5;
 QY 6 NHGSLVAKVAVPLGLCV-----GVFIYVAYIKWHRATATQAFPSITRAAPGARWG--- 56
 Db 28 NHGSL-----LGLCLLQIQTSTGLFLAMHY-----SADATLAFNSIAHLCREVQFGMLL 75
 QY 57 QQAHSPLGTAAAGCHEVY 74
 Db 76 RNLH-----ADGAGMFF 87
 RESULT 15
 Q9T7C8 PRELIMINARY; PRT; 120 AA.
 AC Q9T7C8;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Cytochrome b (Fragment).
 GN CYTB OR CYT B.
 OS Trinomys paratus.
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Hystricognathi; Echimyidae; Trinomys.
 CC NCBI_TaxID=42827;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lara M.C., Patton J.L.;
 RT "Evolutionary diversification of spiny rats (genus Trinomys, Rodentia:
 RL Echimyidae) in the Atlantic Forest of Brazil.";
 RL Zool. J. Linn. Soc. 0:0-0(2000).
 CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
 CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
 CC BOUND TO THE PROTEIN (BY SIMILARITY).
 CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 DR EMBL: AF194294; AAF16096.1; --
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0005746; C:mitochondrial electron transport chain; IEA.
 DR GO: GO:0005739; C:mitochondrion; IEA.
 DR GO: GO:0016491; P:oxidoreductase activity; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.
 DR GO: GO:0006810; P:transport; IEA.
 DR InterPro: IPR005797; Cytb b6 N.
 DR Pfam: PF00033; cytochrome_b_N; 1.
 DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
 KW Electron transport; Heme; Respiratory chain; Transmembrane; Transport;
 KW Mitochondrion.

CC protein or complementary sequences. The polynucleotides are useful for
 CC identifying expressed genes or for physical mapping of human genome. The
 CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight
 CC markers, as a food supplement, for generating antibodies, in medical
 CC imaging, screening and diagnostic assays and for treating cell-
 CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
 CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
 CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
 CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver
 CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
 CC arthritis, etc. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 105 AA;

Query Match 64.7%; Score 414; DB 5; Length 105;
 Best Local Similarity 80.2%; Pred. No. 2.1e-40;
 Matches 77; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 1 MRKISNHSGLRVAKVAYPLGLCVGVPIYVYIKWHRATATQAFSTRAAPGARWQQAH 60
 DB 1 MRKIPNHGTLRWTKVAYPLGLCVGLFIYVYIKWHRASAAQAFSTIAGASGARWTQQAF 60

QY 61 SPLGTAADGHEVPYIGIMFDAGSTGTRVHVQTRPP 96
 DB 61 SPFGSAARGHEVPYIGIMFDAGSTGTRVHVQTRPP 96

RESULT 2

AAB72245
 ID AAB72245 standard; peptide; 14 AA.

XX AAB72245;

XX 14-MAY-2001 (first entry)

XX Peptide used to create anti human CD39-L2 antibodies.
 XX Human CD39-like protein; apyrase; NDPase; platelet function inhibitor;

KW myocardial infarction; cerebral ischaemia; angina; arterial thrombosis;
 KW cerebral artery thrombosis; platelet aggregation; inflammation;
 KW apoptosis; autoimmune disorder; neurological disorder; antibody;
 KW Alzheimer's disease; Parkinson's disease; cancer; CD39-L2.

XX Homo sapiens.

XX WO200110205-A1.

XX 15-FEB-2001.

XX 09-AUG-2000; 2000WO-US021790.

XX 09-AUG-1999; 99US-00370265.

XX 11-JAN-2000; 2000US-00481238.

XX 25-APR-2000; 2000US-00557800.

XX 26-MAY-2000; 2000US-00583231.

XX 30-JUN-2000; 2000US-00608285.

XX (HYSE-) HYSEQ INC.

XX Ford J, Mulero JJ, Yeung G;

XX WPI; 2001-147489/15.

XX Polynucleotides encoding human CD39-like polypeptides, with apyrase
 PT and/or NDPase activity, which are useful in the treatment of pathological
 PT conditions caused by thrombosis (e.g. myocardial infarction) and
 PT inflammatory disorders.
 XX Example 18; Page 102; 203pp; English.

XX This invention relates to polynucleotides encoding human CD39-like

CC polypeptides with apyrase and/or NDPase activity. The polypeptides having
 CC APDase, including NDPase, activity are useful for inhibiting platelet
 CC function and can therefore be used in the prophylaxis or treatment of
 CC pathological conditions caused by or involving thrombosis or excessive
 CC coagulation or excessive platelet aggregation, such as myocardial
 CC infarction, cerebral ischaemia, angina, arterial thrombosis, cerebral
 CC artery thrombosis or intracardiac thrombosis, and conditions associated
 CC with venous thrombosis. CD39-L4 and CD39-L2 polypeptides are useful in
 CC modulating disease states (including platelet aggregation, inflammation
 CC and apoptosis) associated with ADP or other purinergic signalling by
 CC reducing the levels of NDPs. The polypeptides are also useful for
 CC prophylaxis or treatment of inflammation related disorders, such as
 CC disorders involving sepsis or systemic inflammatory response syndrome or
 CC SIRS (and associated conditions such as fever, tachycardia, tachypnea,
 CC cytokine overstimulation); autoimmune disorders such as thrombosis,
 CC atherosclerosis, acute pancreatitis, dermatitis, including psoriasis,
 CC cirrhosis, reperfusion injury, asthma, multiple sclerosis, arthritis;
 CC neurological disorders including neurodegenerative diseases, epilepsy,
 CC depression, Alzheimer's disease, Parkinson's disease, Huntington's
 CC disease, and amyotrophic lateral sclerosis; and cancer. The present
 CC sequence represents a peptide used to create antibodies directed to human
 CC CD39 like protein CD39-L2
 XX
 SQ Sequence 14 AA;

Query Match 12.2%; Score 78; DB 4; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.033;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 TRPPREPTLTTHET 106

DB 1 TRPPREPTLTTHET 14

RESULT 3

AAU66349

ID AAU66349 standard; protein; 95 AA.

XX AAU66349;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #2745.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW urethritis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0199047P.

XX 02-JUN-2000; 2000US-0208841P.

XX 07-JUL-2000; 2000US-0216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX N-PSDB; AAS95729.

XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.

PS Example 1; SEQ ID NO 27544; 1069pp; English

CC Sequences AAU391405-AAU68017 represent Propionibacterium acnes immunogenetic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies can also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences

SQ Sequence 95 AA;

Query Match	11.2%;	Score 72;	DB 4;	Length 95;
Best Local Similarity	34.8%;	Pred. No. 2;		
Matches 23;	Conservative 3;	Mismatches 14;	Indels 26;	Gaps 4

```

QY      36 RATAATGAFFSI--TRAAPGARWGQAAHSPFGTALDGHEVFVFGIMFDAGSTGRVHVHVFQFTR 94
      ||| : ||| ||| ||| ||| :
Db      55 RALGTAAFPRULDRDRPGLRW--DAHSG-GTAPDSHRI----- 89

```

QY	95	PPRETP	100
		:	1
Db	90	PPRESP	95

RESULT 4	
AAU64628	
ID	AAU64628 standard; protein; 95 AA.

AC AAU64628;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #25524

KM SARHO syndrome; synovitis; acne, pustulosis; hyperostosis; osteomyelitis
 KM uveitis; endophthalmitis; bone, joint; central nervous system; ELISA;
 KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KM dermatological; osteopathic; neuroprotectant.

05 *Propionibacterium acnes*.

PN WO200181581-A2.

PD 01-NOV-2001

PF 20-APR-2001; 2001WO-US012865.

PR 21-APR-2000; 2000US-0199047P.

PR 07-JUL-2000; 2000US-0216747P.

PA (CORI-) CORIXA CORP.

PI Skeiky YAW, Persing

Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A,

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

DR WPI; 2001-616774/71

XX

PT Propionibacterium acnes polypeptides and nucleic acids useful for PT vaccinating against and diagnosing infections, especially useful for PT treating acne vulgaris.

PS Example 1; SEQ ID NO 25823; 1069pp; English.

CC Sequences AAU319105-AAU68017 represent *Propionibacterium acnes* immunogen
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC *P. acnes* is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of *P. acnes* in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for *P. acnes* proteins. These antibodies can be used to
CC downregulate expression and activity of *P. acnes* polypeptides and
CC therefore treat *P. acnes* infections. The antibodies may also be used as
CC diagnostic agents for determining *P. acnes* presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC http://wipo.int/pub/published_pct_sequences

Sequence 95 AA;

Query Match	11.2%	Score 72:	DB 4:	Length 95:
Best Local Similarity	34.8%	Pred. No. 2:		
Matches 23, Conservative	3:	Mismatches 14:	Indels 26:	Gaps 4

```

Oy      36  RATAQAFFSI--TRAAPGARWGQAHSPLGTAADGHEVYVYIMEDAGSGTGRVHVQFTR  94
          |||  : ||| ||| ||| ||| :
Db      55  RALGTAAPPRDURDRPGLRW--DAHSG-GTAPDSHRI-----  89

```

QY	95	PPRETP	10
		:	
Db	90	PPRESP	95

RESULT 5	
ABM61147	
ID	ABM61147 standard; protein; 95 AA

AC ABM61147,

DT 20-OCT-2003 (first entry)

DE *Propionibacterium acnes* Predicted ORF-encoded polypeptide #25823.

KW Acne vulgaris; antiseborrheic; dermatolog
KW immunostimulant; immune response; vaccine.

05 *Propionibacterium acnes*.

PN WO2003033515-A1

PD 24-APR-2003

11-OCT-2002; 2002WO-US032727.

PR 15-OCT-2001; 2001US-00978825.

PA (CORI-) CORIXA CORP.

PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;

PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D,

XX

DR N-PSDB; ACF64575.

XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
PS Example 1; SEQ ID NO 25923; 1481pp; English.
XX
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a polypeptide predicted to be encoded by an ORF (open
CC reading frame) contained within the P. acnes polynucleotides of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 95 AA;
Query Match 11.2%; Score 72; DB 6; Length 95;
Best Local Similarity 34.8%; Pred. No. 2;
Matches 23; Conservative 3; Mismatches 14; Indels 26; Gaps 4;
Qy 36 RATATQAFPSI-TRAAPGARWGOQHSPGLTADGHEVFYGVIMFDAGSTGTRVHVQFTR 94
Db 55 RALGTAAPFLDLRLDRPGLRW--DAHSG-GTAPDSHRI----- 89
Qy 95 PPRETP 100
Db 90 PPRESP 95
RESULT 6
ABM62868
XX ABM62868 standard; protein; 95 AA.
AC ABM62868;
XX
DT 20-OCT-2003 (first entry)
XX
DE Propionibacterium acnes predicted ORF-encoded polypeptide #27544.
XX
KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine.
XX
OS Propionibacterium acnes.
XX
PN WO2003033515-A1.
XX
PD 24-APR-2003.
XX
PF 11-OCT-2002; 2002WO-US032727.
XX
PR 15-OCT-2001; 2001US-00978825.
XX

PA (CORI-) CORIXA CORP.
XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI Barth B, Valliave-Douglass J;
XX WPI; 2003-381789/36.
DR N-PSDB; ACF64658.
XX
PS New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
PS Example 1; SEQ ID NO 27544; 1481pp; English.
XX
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a polypeptide predicted to be encoded by an ORF (open
CC reading frame) contained within the P. acnes polynucleotides of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 95 AA;
Query Match 11.2%; Score 72; DB 6; Length 95;
Best Local Similarity 34.8%; Pred. No. 2;
Matches 23; Conservative 3; Mismatches 14; Indels 26; Gaps 4;
Qy 36 RATATQAFPSI-TRAAPGARWGOQHSPGLTADGHEVFYGVIMFDAGSTGTRVHVQFTR 94
Db 55 RALGTAAPFLDLRLDRPGLRW--DAHSG-GTAPDSHRI----- 89
Qy 95 PPRETP 100
Db 90 PPRESP 95
RESULT 7
AAG00192
ID AAG00192 standard; protein; 94 AA.
XX
AC AAG00192;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 4273.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX


```

XX      EPI033401-A2.
XX      06-SEP-2000.
PD
PF      21-FEB-2000; 2000EP-00200610.
XX
PR      26-FEB-1999;   99US-0122487P.
XX
PA      (GEST ) GENSET.
XX
FI      Dumas Milne Edwards J, Duclert A, Giordano J;
DR      MPI; 2000-500381/45.
XX      N-P5DB; AAC00198.
PT       New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PP      obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT      diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS      Claim 13; SEQ ID NO 4273; 71pp + Sequence Listing; English.
XX
CC      The present sequence is a polypeptide encoded by one of a large number of
CC      5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
CC      prepared from total human RNAs or polyA+ RNAs derived from 30 different
CC      tissues. EST sequences usually correspond mainly to the 3' untranslated
CC      region (UTR) of the mRNA because they are often obtained from oligo-dT
CC      primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC      sequences derived from the 5' ends of mRNAs and even in those cases where
CC      longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC      included. 5' ESTs are derived from mRNAs with intact 5' ends and can
CC      therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC      are also used in diagnostic, forensic, gene therapy and chromosome
CC      mapping procedures. They are used to obtain upstream regulatory sequences
CC      and to design expression and secretion vectors
XX
SQ      Sequence 94 AA;
OY      Query Match          9.8%; Score 63; DB 3; Length 94;
        Best Local Similarity 29.2%; Pred.No. 22;
        Matches    14; Conservative    12; Mismatches     20; Indels     2; Gaps     1;
Db      72 VFYGIMPAGSGTGTHVVEQFTRPPEPLTLH--ETPKAVKPGGSAY 117
         |||:|||||::|::::|:|||||:
        47 VKYGVLDNGSHSTLYIKWAKXENDTVGHVQBECRVKGPIGLSKP  94
RESULT 8
ID      ABB78931
XX      ABB78931 standard; peptide; 92 AA.
XX
AC      ABB78931;
DT      30-JUL-2002 (first entry)
DE
DS      Tumour necrosis factor-alpha binding amino acid sequence T14.08.
XX
KW      Protein scaffold; antibody; binding protein; immunoglobulin;
KM      tumour necrosis factor alpha; TNF-alpha; protein framework.
OS      Homo sapiens.
OS      Synthetic.
XX
PN      WO200232925-A2.
XX
PD      25-APR-2002.
PF      16-OCT-2001; 2001WO-US032233.
XX
PR      16-OCT-2000; 2000US-00688566.
XX
PA      (PHYL-) PHYLIS INC.
XX
PI      Lipovsek D, Wagner RW, Kuimelis RG;
```

```

XX DB WP1: 2002-444238/47.
XX
XX PT New non-antibody proteins having an immunoglobulin fold, useful in
XX research, therapeutic or diagnostic fields, particularly as scaffolds for
XX designing proteins with specific properties, e.g. for binding any antigen
XX of interest.
XX
XX PS Claim 47; Fig 25; 94pp; English.
XX
XX The present invention describes a non-antibody protein, comprising a
XX domain having an immunoglobulin-like fold, derived from a reference
XX protein having a mutated amino acid sequence, where the non-antibody
XX protein binds with a Kd at least as tight as 10 nM to a compound that is
XX not bound as tightly by the reference protein. The non-antibody protein
XX is useful as scaffolds for selecting or designing a protein framework
XX with specific and favourable properties, e.g. for binding any antigen of
XX interest, or for deactivating or inactivating antibody molecules. The non-
XX antibody protein is also useful in all areas where antibodies are used,
XX e.g. research, therapeutic or diagnostic fields, and for screening novel
XX binding proteins useful in the above-mentioned fields. The present
XX proteins have thermodynamic properties superior to those of natural
XX antibodies, and can be evolved rapidly in vitro. The present proteins or
XX antibody mimics exhibit improved biophysical properties, such as
XX stability under reducing conditions and solubility at high
XX concentrations. In addition, these molecules are readily expressed and
XX folded in prokaryotic systems (e.g. Escherichia coli), in eukaryotic
XX systems (e.g. yeast), or in in vitro translation systems (e.g. rabbit
XX reticulocyte lysate system). Furthermore, these proteins are extremely
XX amenable to affinity maturation techniques involving multiple cycles of
XX selection, e.g. in vitro selection using RNA-protein fusion technology,
XX phage display or yeast display systems. The present sequence is used in
XX the exemplification of the present invention
XX
XX SQ Sequence 92 AA;
XX
XX Query Match 9.7%; Score 62; DB 5; Length 92;
XX Best Local Similarity 28.2%; Pred. No. 28;
XX Matches 22; Conservative 8; Mismatches 24; Indels 24; Gaps 4;
XX
XX QY 37 ATATGAFFSITRAPAGRWGQAHSPLGTADGHEVYGIIMPAGSSTGRVHVFOETRRP 96
XX 13 ATRPSKILS-----KXPRSH-----HDKRYRILTY--GEIGNSPVGELFVFP 52
XX QY 97 REPTLTHERPKAVKPEL 114
XX DB 53 WGITA---TIDGKPKGV 66
XX
XX RESULT 9
XX ID AU066429 standard; protein; 90 AA.
XX AC AU066429;
XX DT 27-FEB-2002 (first entry)
XX DE Propionibacterium acnes immunogenic protein #27325.
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.
XX OS Propionibacterium acnes.
XX PN W0200181581-A2.
XX PD 01-NOV-2001.
XX PF 20-APR-2001; 2001W0-US012865.
XX PR 21-APR-2000; 2000US-0199047P.

```

PR 02-JUN-2000; 2000US-0208841P.
PR 07-JUL-2000; 2000US-0216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'Maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI; 2001-616774/71.
XX N-PSDB; AAS59733.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
XX Example 1; SEQ ID NO 27624; 1069pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 90 AA;
SQ

Query Match 9.6%; Score 61.5; DB 4; Length 90;
Best Local Similarity 35.2%; Pred. No. 31;
Matches 19; Conservative 3; Mismatches 27; Indels 5; Gaps 2;
QY 36 RATATQAFPSITRAAPGARWQQQAHSPLGTAADGHEVFYIGIMFDAGTGTRVHV 89
DB 41 RMTSTLSLASVDRRLP---WSAHPQPCPRHAADQHE--EPQCAQLGKLGTRPHV 89

RESULT 10
ABM62948
XX ABM62948 standard; protein; 90 AA.
XX
XX AC ABM62948;
XX
XX DT 20-OCT-2003 (first entry)
XX
XX DE Propionibacterium acnes predicted ORF-encoded polypeptide #27624.
XX
XX Acne vulgaris; antisepborrheic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine.
XX
XX OS Propionibacterium acnes.
XX
XX PN WO2003033515-A1.
XX
XX PD 24-APR-2003.
XX
XX PF 11-OCT-2002; 2002WO-US032727.
XX
XX PR 15-OCT-2001; 2001US-00978825.
XX
XX PA (CORI-) CORIXA CORP.

XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI Barth B, Vallieue-Douglas J;
XX
XX WPI; 2003-381789/36.
XX N-PSDB; ACF64662.
XX
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
XX Example 1; SEQ ID NO 27624; 1481pp; English.
XX
XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a polypeptide predicted to be encoded by an ORF (open
CC reading frame) contained within the P. acnes polynucleotides of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 90 AA;
SQ

Query Match 9.6%; Score 61.5; DB 6; Length 90;
Best Local Similarity 35.2%; Pred. No. 31;
Matches 19; Conservative 3; Mismatches 27; Indels 5; Gaps 2;
QY 36 RATATQAFPSITRAAPGARWQQQAHSPLGTAADGHEVFYIGIMFDAGTGTRVHV 89
DB 41 RMTSTLSLASVDRRLP---WSAHPQPCPRHAADQHE--EPQCAQLGKLGTRPHV 89

RESULT 11
ADD26584
XX ADD26584 standard; protein; 100 AA.
XX
XX AC ADD26584;
XX
XX DT 15-JAN-2004 (first entry)
XX
XX DE BGS-19 immunoglobulin-like domain 1 SEQ ID NO:9.
XX
XX human; cell surface protein; immunoglobulin; BGS-19; cytostatic;
KW gynaecological; immunosuppressive; antiinflammatory; antiasthmatic;
KW antidiabetic; dermatological; gene therapy.
XX
XX OS Homo sapiens.
XX
XX PN WO2003083078-A2.
XX
XX PD 09-OCT-2003.
XX

PF 28-MAR-2003; 2003WO-US009676.
 XX
 PR 28-MAR-2002; 2002US-0368422P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Lee LM, Feder JN, Siemers NO, Wu S, Chen J;
 XX WPI; 2003-604052/75.
 DR
 PT New isolated nucleic acid molecule encoding BGS-19 polypeptides, useful
 PT for preventing, treating or ameliorating a medical condition, such as a
 PT disorder related to aberrant immunoglobulin cell surface receptor
 PT activity.
 XX
 PS Disclosure; SEQ ID NO 9; 224pp; English.
 CC The present invention describes human cell surface protein with
 CC immunoglobulin folds, designated BGS-19 (I). (I) has cytostatic,
 CC synaescological, immunosuppressive, antiinflammatory, antiasthmatic,
 CC antidiabetic and dermatological activities, and can be used in gene
 CC therapy. (I) can be used for preventing, treating or ameliorating a
 CC medical condition, such as a disorder related to aberrant immunoglobulin
 CC cell surface receptor activity; a cellular adhesion disorder; a disorder
 CC related to hyper- or hypo-immunoglobulin receptor activity; a disorder
 CC related to aberrant signal transduction; a reproductive disorder; a
 CC female reproductive disorder; an ovarian disorder; ovarian cancer;
 CC dysfunctional uterine bleeding; amenorrhoea; primary dysmenorrhoea;
 CC sexual dysfunction; infertility; pelvic inflammatory disease;
 CC endometriosis; placental aromatase deficiency; premature menopause;
 CC placental dysfunction; hormone deficiency; oestrogen deficiency; aberrant
 CC androgen metabolism; aberrant onset of female puberty; aberrant showing
 CC of female primary sexual characteristics; aberrant showing of female
 CC secondary sexual characteristics; precocious puberty; precocious
 CC pseudopuberty; incomplete isosexual precocity; premature thelarche;
 CC premature adrenarche; premature pubarche; polycystic ovarian disease;
 CC aberrant ovarian cycle; menorrhagia; metrorrhagia; menometrorrhagia;
 CC dysmenorrhoea; hypomenorrhoea; polymenorrhoea; dysfunctional uterine
 CC bleeding; resistant-ovary syndrome; hermaphroditism; immune disorders;
 CC inflammatory disorders; arthritis; asthma; immunodeficiency diseases such
 CC as AIDS; leukaemia; rheumatoid arthritis; granulomatous diseases;
 CC inflammatory bowel disease; sepsis; acne; neutropenia; neutrophilia;
 CC psoriasis; hypersensitivities; such as T-cell mediated cytotoxicity;
 CC immune reactions to transplanted organs and tissues; such as host-versus-
 CC graft and graft-versus-host diseases; or autoimmunity disorders;
 CC autoimmune infertility; Addison's Disease; haemolytic anaemia;
 CC antiphospholipid syndrome; rheumatoid arthritis; dermatitis; allergic
 CC encephalomyelitis; glomerulonephritis; Goodpasture's Syndrome; Graves'
 CC Disease; Multiple Sclerosis; Myasthenia Gravis; Neuritis; Ophthalmia;
 CC Bullous Pemphigoid; Pemphigus; Polyendocrinopathies; Purpura; Reiter's
 CC Disease; Stiff-Man Syndrome; autoimmune thyroiditis; Systemic Lupus
 CC Erythematosus; Autoimmune Pulmonary Inflammation; Guillain-Barre Syndrome
 CC ; Insulin dependent diabetes mellitus; autoimmune inflammation eye
 CC disease; lens tissue injury; demyelination; systemic lupus erythematosus;
 CC drug induced haemolytic anaemia; rheumatoid arthritis; Sjogren's disease;
 CC and scleroderma. The present sequence is used in the exemplification of
 CC the present invention.
 CC
 XX
 SO Sequence 100 AA;
 Query Match 9.5%; Score 60.5; DB 7; Length 100;
 Best Local Similarity 29.7%; Pred. No. 47;
 Matches 27; Conservative 12; Mismatches 33; Indels 19; Gaps 5;
 QY 9 SLRYAK-VAAPDLGCVGVFTIVAYIK--WRATPQAQFFSITRAAPARWGQQAHSPLG 65
 DB 10 SLQGRQVPYPRFGICVVCVSLSTPRDGMDESTRAYGWKRTSP-----KTGAPVAT 63
 QY 66 AADGHEVFGIMPAGSTGRVHVFOPTRP 96
 DB 64 NNQSRV-----EMSTRDR---FOLTGP 84

RESULT 12
 ID AU31171
 XX AU31171 standard; protein; 114 AA.
 XX
 AC AU31171;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #1662.
 XX
 KM Human; vaccination; gene therapy; nutritional supplement;
 KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US008656.
 XX
 PR 18-APR-2000; 2000US-00552929.
 PR 26-JAN-2001; 2001US-00770160.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX WPI; 2001-611725/70.
 DR
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.
 XX
 PS Claim 20; Page 413; 765pp; English.
 XX
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukemias.
 CC AAU93510-AU93304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention
 CC
 XX
 SO Sequence 114 AA;
 Query Match 9.1%; Score 58.5; DB 4; Length 114;
 Best Local Similarity 27.0%; Pred. No. 95;
 Matches 30; Conservative 10; Mismatches 22; Indels 49; Gaps 7;
 QY 30 AYIKWRATPQAQFFSITRAAPARWGQQAHSPLGTAADGHEVFGIWDAGSTGRVH- 88
 DB 19 AGVCWHDRLRSIOA-----PPGSR-----APASRV-----AGTTGTRHHA 56
 QY 89 -VPOF-----TRPREPTLTHETFAVYRGSL 115
 DB 57 RLIFFFFFFFFLVETGFHAYSEDDDLFTSGDLPIFLTSGS--AGLPGMS 105
 RESULT 13
 ID ABB40819
 XX ABB40819 standard; peptide; 82 AA.

```
AC ABB40819;
XX 04-FEB-2002 (first entry)
XX Peptide #8325 encoded by human foetal liver single exon probe.
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX Homo sapiens.
XX WO200157277-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000669.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.
XX Claim 27; SEQ ID NO 33454; 639pp + Sequence Listing; English.
XX The invention relates to a single exon nucleic acid probe for measuring
XX single exon gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human fetal liver. The
XX present sequence is a peptide encoded by a single exon nucleic acid probe
XX of the invention. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pat_sequences
XX Sequence 82 AA;
XX
XX Query Match 9.1%; Score 58; DB 4; Length 82;
XX Best Local Similarity 30.9%; Pred. No. 71;
XX Matches 21; Conservative 9; Mismatches 18; Indels 20; Gaps 4;
XX
XX QY 40 TQAFSSI-TRAAPCARWG-QQAHSPLGTAADGHEVFYGMFDAGSTGTRVHVQFTRPPR 97
XX ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
XX 17 TQGLGIGQKSAPRAPWALKVQGPLEPQA-----PKVN-RETTTRPPR 58
XX
XX QY 98 ETPTLTHE 105
XX :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
XX 59 KSPSLPOE 66
XX
XX RESULT 14
XX AAM34585
XX ID AAM34585 standard; protein; 82 AA.
XX
XX AC AAM34585;
XX
XX 17-OCT-2001 (first entry)
XX
XX Peptide #8622 encoded by probe for measuring placental gene expression.
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX Genetic disorder.
XX Homo sapiens.
XX
```

```
XX WO200157272-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000663.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX Claim 27; SEQ ID NO 34854; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP:
XX see AAI31315-AAJ57546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX Sequence 82 AA;
XX
XX Query Match 9.1%; Score 58; DB 4; Length 82;
XX Best Local Similarity 30.9%; Pred. No. 71;
XX Matches 21; Conservative 9; Mismatches 18; Indels 20; Gaps 4;
XX
XX QY 40 TQAFSSI-TRAAPCARWG-QQAHSPLGTAADGHEVFYGMFDAGSTGTRVHVQFTRPPR 97
XX ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
XX 17 TQGLGIGQKSAPRAPWALKVQGPLEPQA-----PKVN-RETTTRPPR 58
XX
XX QY 98 ETPTLTHE 105
XX :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
XX 59 KSPSLPOE 66
XX
XX RESULT 15
XX AAM74471
XX ID AAM74471 standard; protein; 82 AA.
XX
XX AC AAM74471;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 34777.
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000668.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX
```

PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.

PS Example 4; SEQ ID NO 34777; 658bp + Sequence listing; English.

CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention

XX
SQ Sequence 82 AA;

Query Match 9.1%; Score 58; DB 4; Length 82;
Best Local Similarity 30.9%; Pred. No. 71;
Matches 21; Conservative 9; Mismatches 18; Indels 20; Gaps 4;

OY 40 TOAFPSI-TRAAPGRNG-QQHSPLGTADGHEVFGIMFDAGSTGTRVHVFQTRPPR 97
Db 17 TGGFLGIGKSAAPRAFWALKVAGCPLEPQA-----PKVN-RETRRPPR 58

OY 98 EPTLVTHE 105
Db 59 KSPSLPQE 66

Search completed: June 8, 2004, 10:00:18
Job time : 57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 10:01:46 ; Search time 42 Seconds
(without alignments)
803,825 Million cell updates/sec

Title: US-09-905-589a-2_COPY_1_120

Perfect score: 640

Sequence: 1 MRKSNHGSRLKAVKAYPLG.....TLTHETPKAYKPLSLAYADD 120

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 526131

Minimum DB seq length: 0
Maximum DB seq length: 120

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	12.2	14	US-10-286-926-44	Sequence 44, Appl
2	78	12.2	108	US-10-424-599-217882	Sequence 217882,
3	66.5	10.4	119	US-10-424-599-219167	Sequence 219167,
4	66	10.3	80	US-10-062-831-220	Sequence 220, App
5	65	10.3	80	US-10-062-831-220	Sequence 220, App
6	63	9.8	61	US-10-424-599-282823	Sequence 282823,
7	60.5	9.5	100	US-10-403-938-9	Sequence 9, Appl
8	59	9.2	81	US-10-424-599-230474	Sequence 230474,
9	58	9.1	82	US-09-864-761-44698	Sequence 44698, A
10	58	9.1	91	US-10-424-599-154715	Sequence 154715,
11	57.5	9.0	112	US-10-264-237-1834	Sequence 1834, Ap
12	57.5	9.0	118	US-10-108-2604-2613	Sequence 2613, Ap
13	57	8.9	63	US-10-424-599-168571	Sequence 168571,
14	57	8.9	72	US-09-864-761-33368	Sequence 33368, A
15	57	8.9	108	US-09-738-626-5443	Sequence 5443, Ap

16	56.5	8.8	81	US-10-424-599-255385	Sequence 255385,
17	56.5	8.8	87	US-10-424-599-172253	Sequence 172253,
18	56.5	8.8	119	US-10-425-114-57544	Sequence 57544, A
19	56	8.7	52	US-10-424-599-181841	Sequence 181841,
20	55.5	8.7	58	US-09-835-147-31	Sequence 31, Appl
21	55	8.6	101	US-10-108-2604-4537	Sequence 4537, Ap
22	54.5	8.5	61	US-10-424-599-160185	Sequence 160185,
23	54.5	8.5	89	US-10-424-599-218455	Sequence 218455,
24	54	8.4	92	US-10-424-599-18182	Sequence 18182,
25	53.5	8.4	70	US-09-867-550-364	Sequence 364, App
26	53.5	8.4	15	US-10-264-049-2513	Sequence 2513, Ap
27	53	8.3	82	US-10-424-599-201768	Sequence 201768,
28	53	8.3	107	US-10-104-047-3230	Sequence 3230, Ap
29	53	8.3	112	US-10-424-599-282912	Sequence 282912,
30	52.5	8.2	54	US-10-083-357-1261	Sequence 1261, Ap
31	52.5	8.2	64	US-10-424-599-250977	Sequence 250977,
32	52.5	8.2	86	US-10-001-885-166	Sequence 166, App
33	52.5	8.2	96	US-10-302-456-1	Sequence 1,
34	52.5	8.2	116	US-10-282-1522A-62053	Sequence 62053, A
35	52.5	8.2	117	US-10-424-599-192570	Sequence 192570,
36	52	8.1	78	US-10-424-599-236156	Sequence 236156,
37	52	8.1	98	US-10-424-599-239102	Sequence 239102,
38	52	8.1	103	US-10-424-599-279416	Sequence 279416,
39	52	8.1	110	US-10-029-386-29859	Sequence 29859, A
40	51.5	8.0	73	US-10-424-599-274655	Sequence 274655,
41	51.5	8.0	82	US-09-864-761-39751	Sequence 39751, A
42	51	8.0	9	US-10-424-599-229454	Sequence 229454,
43	51	8.0	73	US-10-424-599-171425	Sequence 171425,
44	51	8.0	87	US-10-424-599-277082	Sequence 277082,
45	51	8.0	113	US-10-424-599-274427	Sequence 274427,

ALIGNMENTS

RESULT 1
US-10-286-926-44 Application US/10286926
Sequence 44, Application US/10286926
Publication No. US20030175752A1
GENERAL INFORMATION:
APPLICANT: Ford, John
APPLICANT: Maleiro, Julio
TITLE OF INVENTION: Methods and Materials Relating to CD39-Like
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 28110/36457CON
CURRENT APPLICATION NUMBER: US/10/266, 926
CURRENT FILING DATE: 2002-11-01
PRIOR APPLICATION NUMBER: 09/557, 800
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/481, 238
PRIOR FILING DATE: 2000-01-11
PRIOR APPLICATION NUMBER: 09/370, 265
PRIOR FILING DATE: 1999-08-09
PRIOR APPLICATION NUMBER: PCT/US99/16180
PRIOR FILING DATE: 1999-07-16
PRIOR APPLICATION NUMBER: 09/350836
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: 09/273447
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: 09/122449
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: 09/244444
PRIOR FILING DATE: 1999-02-04
PRIOR APPLICATION NUMBER: 09/118, 205
PRIOR FILING DATE: 1998-07-16
NUMBER OF SEQ ID NOS: 54
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 44
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: peptide
US-10-286-926-44

Query Match 12.2%; Score 78; DB 14; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 TRPPRETPLTHET 106
Db 1 TRPPRETPLTHET 14

RESULT 2

US-10-424-599-217882
; Sequence 217882, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 217882
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_38774C.1.pap
US-10-424-599-217882

Query Match 12.2%; Score 78; DB 12; Length 108;
Best Local Similarity 68.4%; Pred. No. 0.15;
Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 74 YGIMFDAGSTGRVHFQF 92
Db 68 YAVIFDAGSGSRVHFHF 86

RESULT 3

US-10-424-599-219167
; Sequence 219167, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 219167
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(119)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_39935C.1.pap
US-10-424-599-219167

Query Match 10.4%; Score 66.5; DB 12; Length 119;
Best Local Similarity 37.0%; Pred. No. 4.6;

Matches 17; Conservative 5; Mismatches 23; Indels 1; Gaps 1;
QY 76 IMFADAGSTGRVHFQFTRPPRETPT-LTHETPKAVKPGLSAYADD 120
Db 28 VIFDGGTGPRVHFHQAQNDLLPMGVGLNKRITPGFXAYGVD 73

RESULT 4

US-10-062-831-220
; Sequence 220, Application US/10062831
; Publication No. US20030105297A1
; GENERAL INFORMATION:
; APPLICANT: Steven M. Ruben, et al.
; TITLE OF INVENTION: 32 Human Secreted Proteins
; FILE REFERENCE: P2006P1
; CURRENT APPLICATION NUMBER: US/10/062,831
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: 09/690,454
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: PCT/US98/10868
; PRIOR FILING DATE: May 28, 1998
; PRIOR APPLICATION NUMBER: 60/044,039
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,093
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,190
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/050,935
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,101
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,356
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/056,250
; PRIOR FILING DATE: August 29, 1997
; PRIOR APPLICATION NUMBER: 60/056,296
; PRIOR FILING DATE: August 29, 1997
; PRIOR APPLICATION NUMBER: 60/056,293
; PRIOR FILING DATE: August 29, 1997
; NUMBER OF SEQ ID NOS: 229
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 220
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (72)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-062-831-220

Query Match 10.3%; Score 66; DB 14; Length 80;
Best Local Similarity 57.9%; Pred. No. 3.2;
Matches 11; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 51 PGARWGQQAHSPLGTAADG 69
Db 40 PRIRWGQEAHVPAQAQEG 58

RESULT 5

US-10-062-599-220
; Sequence 220, Application US/10062599
; Publication No. US20030195346A1
; GENERAL INFORMATION:
; APPLICANT: Steven M. Ruben, et al.
; TITLE OF INVENTION: 32 Human Secreted Proteins
; FILE REFERENCE: P2006P1
; CURRENT APPLICATION NUMBER: US/10/062,599
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: 09/690,454
; PRIOR FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: 09/189,144

PRIOR FILING DATE: 1998-11-10
PRIOR APPLICATION NUMBER: 60/044,039
PRIOR FILING DATE: May 30, 1997
PRIOR APPLICATION NUMBER: 60/048,093
PRIOR FILING DATE: May 30, 1997
PRIOR APPLICATION NUMBER: 60/048,190
PRIOR FILING DATE: May 30, 1997
PRIOR APPLICATION NUMBER: 60/050,935
PRIOR FILING DATE: May 30, 1997
PRIOR APPLICATION NUMBER: 60/048,101
PRIOR FILING DATE: May 30, 1997
PRIOR APPLICATION NUMBER: 60/048,356
PRIOR FILING DATE: May 30, 1997
PRIOR APPLICATION NUMBER: 60/056,250
PRIOR FILING DATE: August 29, 1997
PRIOR APPLICATION NUMBER: 60/056,296
PRIOR FILING DATE: August 29, 1997
PRIOR APPLICATION NUMBER: 60/056,293
PRIOR FILING DATE: August 29, 1997
NUMBER OF SEQ ID NOS: 229
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 220
LENGTH: 80
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (72)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-062-599-220

Query Match 10.3%; Score 66; DB 14; Length 80;
Best Local Similarity 57.9%; Pred. No. 3.2;
Matches 11; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 51 PGARWGOAHSPLGTADG 69
DB 40 PRIRWGGAHVPAAGAOEG 58

RESULT 6
US-10-424-599-282823

Sequence 282823, Application US/10424599
Publication No. US20040031072A1

GENERAL INFORMATION:

APPLICANT: La Rosa Thomas J

APPLICANT: Kovalic David K

APPLICANT: Zhou Yihua

APPLICANT: Cao Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

FILE REFERENCE: 38-21(53223)B

CURRENT APPLICATION NUMBER: US/10/424,599

CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 285684

SEQ ID NO 282823

LENGTH: 61

TYPE: PRT

ORGANISM: Glycine max

FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_97411C.1.pep

US-10-424-599-282823

Query Match 9.8%; Score 63; DB 12; Length 61;
Best Local Similarity 29.6%; Pred. No. 5.3;
Matches 21; Conservative 8; Mismatches 30; Indels 12; Gaps 3;

OY 28 YVAIYKMRATATQAFESITRAAPGARWGOAHSPLGTADGHEVFYGIWFDAGSTGTRV 87
DB 2 YLVVYS--DGDTPDFTSAVVVSP-----FPL-TPSAGNKSXYALALASGSGATY 49

OY 88 HVFOFTREPRE 98
|||

DB 50 GINGVNPBRE 60

RESULT 7

US-10-403-938-9

Sequence 9, Application US/10403938

Publication No. US20040025195A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: NOVEL HUMAN CELL SURFACE PROTEIN WITH IMMUNOGLOBULIN FOLDS,

FILE REFERENCE: D0227 NP

CURRENT APPLICATION NUMBER: US/10/403,938

CURRENT FILING DATE: 2003-03-28

PRIOR APPLICATION NUMBER: U.S. 60/368,422

PRIOR FILING DATE: 2002-03-28

NUMBER OF SEQ ID NOS: 88

SOFTWARE: Patent In version 3.2

SEQ ID NO 9

LENGTH: 100

TYPE: PRT

ORGANISM: Homo sapiens

US-10-403-938-9

Query Match 9.5%; Score 60.5; DB 16; Length 100;
Best Local Similarity 29.7%; Pred. No. 20;
Matches 27; Conservative 12; Mismatches 33; Indels 19; Gaps 5;

OY 9 SLRVAK-VAYPLGLCVGFIVYVYIK--WHRATATQAFESITRAAPGARWGOAHSPLGT 65
DB 10 SLQVQRQVPEPGLCVIVSCLNSYPRDGDDESTAAYGVFKGRTSP-----KTGAPVAT 63

OY 66 AADGHEVFYGIWFDAGSTGTRVHVFOFTRP 96
DB 64 NNQSRREV-----EMSTRDR---FQLTGDP 84

RESULT 8

US-10-424-599-230474

Sequence 230474, Application US/10424599

Publication No. US20040031072A1

GENERAL INFORMATION:

APPLICANT: La Rosa Thomas J

APPLICANT: Kovalic David K

APPLICANT: Zhou Yihua

APPLICANT: Cao Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

FILE REFERENCE: 38-21(53223)B

CURRENT APPLICATION NUMBER: US/10/424,599

CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 285684

SEQ ID NO 230474

LENGTH: 81

TYPE: PRT

ORGANISM: Glycine max

FEATURE:

OTHER INFORMATION: Clone ID: PAT_MRT3847_5013C.1.pep

US-10-424-599-230474

Query Match 9.2%; Score 59; DB 12; Length 81;
Best Local Similarity 45.0%; Pred. No. 24;
Matches 9; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

OY 73 FYGIMFDAGSTGTRVHVFOF 92
DB 21 YTTVVDDCGGTGTRVAVVEM 40

RESULT 9

US-09-864-761-44698

Sequence 44698, Application US/09864761

Patent No. US20020048763A1

Db 59 KSPSLPQE 66

RESULT 10

US-10-424-599-154715

; Sequence 154715, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)/B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 154715

; LENGTH: 91

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)..(91)

; OTHER INFORMATION: unsure at all xaa locations

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_110729C.1.pep

US-10-424-599-154715

Query Match 9.1%; Score 58; DB 12; Length 91;

Best Local Similarity 30.8%; Pred. No. 37;

Matches 16; Conservative 4; Mismatches 26; Indels 6; Gaps 2;

Qy 37 ATATQAFSTTAAAGARWGQAHSPLGTAAAGCHEVFYGYMFDAGSTGTTRVH 88

Db 23 ANAINMKXYHRMSPGSLWAHGHSP-GQTLDSIEV-----HDVHXTCANTH 68

RESULT 11

US-10-264-237-1834

; Sequence 1834, Application US/10264237

; Publication No. US20040009491A1

; GENERAL INFORMATION:

; APPLICANT: Birse et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PA131P1

; CURRENT APPLICATION NUMBER: US/10/264,237

; CURRENT FILING DATE: 2002-10-04

; PRIOR APPLICATION NUMBER: PCT/US01/16450

; PRIOR FILING DATE: 2001-05-18

; PRIOR APPLICATION NUMBER: US 60/205,515

; PRIOR FILING DATE: 2000-05-19

; NUMBER OF SEQ ID NOS: 2876

; SOFTWARE: PatentIn Ver. 3.1

; SEQ ID NO 1834

; LENGTH: 112

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-264-237-1834

Query Match 9.0%; Score 57.5; DB 15; Length 112;

Best Local Similarity 25.0%; Pred. No. 55;

Matches 20; Conservative 17; Mismatches 20; Indels 23; Gaps 5;

Qy 8 GSLRVAKVAVPLGLCV----GVFIYVAYIKWHRATATQAFFSITRAAPGARWGQAHSPL 63

Db 45 GNLALA-ILLPLGLVIVLGSVVYVYTKLQ-----GKSLFGFS-----GSHSYSPI 89

Qy 64 GTAADGHEVFGIMFDAGST 83

Db 90 TVESD-----FSNPLYEAGDT 105

Db 59 KSPSLPQE 66

RESULT 10

US-10-424-599-154715

; Sequence 154715, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)/B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 154715

; LENGTH: 91

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)..(91)

; OTHER INFORMATION: unsure at all xaa locations

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_110729C.1.pep

US-10-424-599-154715

Query Match 9.1%; Score 58; DB 12; Length 91;

Best Local Similarity 30.8%; Pred. No. 37;

Matches 16; Conservative 4; Mismatches 26; Indels 6; Gaps 2;

Qy 37 ATATQAFSTTAAAGARWGQAHSPLGTAAAGCHEVYFVYGMFDAGSTGTRVH 88

Db 23 ANAINMKYVHRMSPGSLWAHGHSP-GQTLDSIEV-----HDVHXTCANTH 68

RESULT 11

US-10-264-237-1834

; Sequence 1834, Application US/10264237

; Publication No. US20040009491A1

; GENERAL INFORMATION:

; APPLICANT: Birse et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PA131P1

; CURRENT APPLICATION NUMBER: US/10/264,237

; CURRENT FILING DATE: 2002-10-04

; PRIOR APPLICATION NUMBER: PCT/US01/16450

; PRIOR FILING DATE: 2001-05-18

; PRIOR APPLICATION NUMBER: US 60/205,515

; PRIOR FILING DATE: 2000-05-19

; NUMBER OF SEQ ID NOS: 2876

; SOFTWARE: PatentIn Ver. 3.1

; SEQ ID NO 1834

; LENGTH: 112

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-264-237-1834

Query Match 9.0%; Score 57.5; DB 15; Length 112;

Best Local Similarity 25.0%; Pred. No. 55;

Matches 20; Conservative 17; Mismatches 20; Indels 23; Gaps 5;

Qy 8 GSLRVAKVAVPLGLCV----GVFIYVAYIKWHRATATQAFFSITRAAPGARWGQAHSPL 63

Db 45 GNLALA-ILLPLGLVIVLGSVYIYTKLQ-----GKSLFGFS-----GSHSYSPI 89

Qy 64 GTAADGHEVYGMFDAGST 83

Db 90 TVESD-----FSNPLYEAGDT 105

RESULT 12
US-10-108-260A-2613
; Sequence 2613, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2613
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-2613

Query Match 9.0%; Score 57.5; DB 15; Length 118;
Best Local Similarity 27.4%; Pred. No. 59;
Matches 32; Conservative 8; Mismatches 40; Indels 37; Gaps 6;

QY 11 RVAKYAVPLGLCV-GVFI-----YVAYIKMHRATATGAFSITR--AAGARWGQQAHS 61
DB 25 RIGYITPLATSIILMFIPLASFLVSYC-----DSIPRTVAYPCSGWALNSK 73
QY 62 PLGTADGHEVYVYGMFDAGSTGTRVHVQFTRPRETPTLTHTFKAVK--GLSA 116
DB 74 WEMQAAGAPSPQVYDLFSGCGP-----ETVPATLPASEPMCGILA 115

RESULT 13
US-10-424-599-165371
; Sequence 165371, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 165371
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(63)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_120346C.1.pep
US-10-424-599-165371

Query Match 8.9%; Score 57; DB 12; Length 63;
Best Local Similarity 29.8%; Pred. No. 31;
Matches 14; Conservative 8; Mismatches 21; Indels 4; Gaps 2;

QY 28 YVAYIKMHRATATQ-AFSITRAAPGARWGQQAHSPLCTADGHEVF 73
DB 20 YKAYFTKHKSKNSQKSPRSLTLTFPGQMW---HMKRCFPGHQSY 63

RESULT 14
US-09-864-761-33368
; Sequence 33368, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmlca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 33368
LENGTH: 72
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:

OTHER INFORMATION: MAP TO AC005973.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 8.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 7.8
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6.6
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.6
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 8.5
US-09-864-761-33368

Query Match 8.9%; Score 57; DB 9; Length 72;
Best Local Similarity 51.9%; Pred. No. 36;
Matches 14; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 37 ATATCAFFSITRAAPGARWGQQAHSPL 63
DB 44 AATTAABSETTAAAPAWGGEAPVPL 70

```

RESULT 15
US-09-738-626-5443
; Sequence 5443, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIALI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5443
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5443

Query Match      8.9%; Score 57; DB 9; Length 108;
Best Local Similarity 28.8%; Pred. No. 61;
Matches 19; Conservative 5; Mismatches 42; Indels 0; Gaps 0;

Qy 53 ARWCCQAHSPIGTAADGHEVEYGYIMFDAGSTGTRVHVVFQTRPPRETPTLTHTETKAVKP 112
Db 20 SRWNKFAVSPIGFRAGALRFLGFSAAASTPSSVALVLMVASPSRSASALIASSSKASVS 79

Qy 113 GLSAYA 118
Db 80 GLYSVA 85

```

Search completed: June 8, 2004, 10:07:25
Job time : 43 secs

Tue Jun 8 12:10:42 2004

us-09-905-589a-2_copy_1_120.ra1

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 09:59:16 ; Search time 22 Seconds
(without alignments)
281.596 Million cell updates/sec

Title: US-09-905-589A-2_COPY_1_120
Perfect score: 640
Sequence: 1 MKRISNHSGLRVAKVAVPLG.....TLTHTFYAKPKGLSAVAD 120

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 277640

Minimum DB seq length: 0
Maximum DB seq length: 120

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_5/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/6CTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1e1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	12.2	14	US-09-608-285A-44	Sequence 44, Appl
2	78	12.2	14	US-09-557-800C-44	Sequence 44, Appl
3	67.5	10.5	96	US-09-252-991A-27819	Sequence 27819, A
4	66	10.3	80	US-09-690-454-220	Sequence 220, App
5	54	8.4	108	US-08-466-033-106	Sequence 106, App
6	54	8.4	108	US-08-444-733-106	Sequence 106, App
7	54	8.4	108	US-08-464-134-106	Sequence 106, App
8	54	8.4	108	US-08-461-361-106	Sequence 106, App
9	54	8.4	108	US-08-485-910-106	Sequence 106, App
10	54	8.4	108	PCT-US95-06266-89	Sequence 89, Appl
11	53	8.0	109	US-09-543-681A-6609	Sequence 6609, Ap
12	51.5	8.0	113	US-09-489-039A-8242	Sequence 8242, Ap
13	51	8.0	101	US-09-489-039A-12049	Sequence 12049, A
14	51	8.0	108	US-08-959-812-110	Sequence 108, Appl
15	50	7.8	90	US-09-621-976-4538	Sequence 4538, Ap
16	50	7.8	90	US-09-621-976-4538	Sequence 206, Appl
17	49	7.7	67	US-09-230-041-26	Sequence 26, Appl
18	49	7.7	70	US-09-079-372-12	Sequence 12, Appl
19	49	7.7	100	US-09-079-372-14	Sequence 14, Appl
20	48.5	7.5	116	US-09-489-039A-11188	Sequence 11188, A
21	48	7.5	52	US-09-621-976-5873	Sequence 5873, Ap
22	48	7.5	96	US-09-369-247-72	Sequence 72, Appl
23	48	7.5	112	US-09-621-976-5677	Sequence 5677, Ap
24	48	7.5	112	US-09-621-976-5805	Sequence 5805, Ap
25	47.5	7.4	40	US-08-469-260A-467	Sequence 467, App
26	47.5	7.4	40	US-08-469-260A-467	Sequence 467, App
27	47.5	7.4	40	US-08-467-344A-467	Sequence 467, App

28	47.5	7.4	83	4	US-09-252-991A-19508	Sequence 19508, A
29	47.5	7.4	87	4	US-09-205-258-894	Sequence 894, Appl
30	47.5	7.4	114	1	US-08-165-754-1	Sequence 1, Appl
31	47.5	7.4	114	1	US-07-845-5925-1	Sequence 1, Appl
32	47	7.3	62	4	US-08-621-976-7044	Sequence 7044, Appl
33	47	7.3	68	4	US-08-107-532A-3872	Sequence 3872, Ap
34	47	7.3	72	4	US-09-540-326-2170	Sequence 2170, Ap
35	47	7.3	89	4	US-09-489-039A-13860	Sequence 13860, A
36	47	7.3	114	4	US-09-252-991A-25775	Sequence 25775, A
37	47	7.3	114	4	US-09-489-039A-9698	Sequence 9698, Ap
38	47	7.3	117	4	US-08-252-991A-31460	Sequence 31460, A
39	46.5	7.3	73	4	US-08-252-991A-20704	Sequence 20704, A
40	46.5	7.3	93	4	US-09-328-352-5221	Sequence 3221, Ap
41	46.5	7.3	94	4	US-09-621-976-5714	Sequence 5714, Ap
42	46.5	7.3	107	4	US-09-621-976-5595	Sequence 5595, Ap
43	46.5	7.3	117	4	US-09-134-000C-3881	Sequence 3881, Ap
44	46	7.2	25	1	US-07-752-101A-6	Sequence 6, Appl
45	46	7.2	33	3	US-07-741-453A-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-608-285A-44
; Sequence 44, Application US/09608285A
; Patent No. 6335013
; GENERAL INFORMATION:
; APPLICANT: Mulero, John
; APPLICANT: Ford, Julio
; APPLICANT: Young, George
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE
; FILE REFERENCE: 28110/36570
; CURRENT APPLICATION NUMBER: US/09/608, 285A
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/583, 231
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 09/557, 800
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/481, 238
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 09/370, 265
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350, 836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273, 447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/244, 444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/122, 449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118, 205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 44
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: peptide
US-09-608-285A-44
Query Match 12.2% Score 78; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
93 TRPREPTLTHET 106
1 TRPREPTLTHET 14

Best Local Similarity 29.1%; Pred. No. 0.54; Mismatches 13; Indels 17; Gaps 4;
Matches 23; Conservative 13; Mismatches 26; Indels 17; Gaps 4;
QY 49 AAPCARWQQAAHSPGLGTAADGHEVFGIMFDAGSTGTRVHVFQFTRPPRTPLTHE--- 105
DB 21 ALPGR--GRISAPLGLATSGSLSHGL-----TRVLTITNGGSEENPLPDQSPG 70
QY 106 --TPKAVK--PGLSAYADD 120
DB 71 DRTYRERREFGCPAGTDD 89
RESULT 4
US-09-690-454-220
; Sequence 220, Application US/09690454
; Patent No. 6531447
; GENERAL INFORMATION:
; APPLICANT: Steven M. Ruben, et al.
; TITLE OF INVENTION: 32 Human Secreted Proteins
; FILE REFERENCE: P2006P1
; CURRENT APPLICATION NUMBER: US/09/690,454
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: 09/189,144
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: 60/044,039
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,093
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,190
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/050,935
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,101
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,356
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/056,250
; PRIOR FILING DATE: August 29, 1997
; PRIOR APPLICATION NUMBER: 60/056,296
; PRIOR FILING DATE: August 29, 1997
; PRIOR APPLICATION NUMBER: 60/056,293
; PRIOR FILING DATE: August 29, 1997
; NUMBER OF SEQ ID NOS: 229
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 220
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (72)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-690-454-220
Query Match 10.3%; Score 66; DB 4; Length 80;
Best Local Similarity 57.9%; Pred. No. 0.56; Mismatches 2; Indels 0; Gaps 0;
Matches 11; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 51 PGARWQQAAHSPGLGTAADG 69
DB 40 PRIRWQEAHVPAQAQEG 58
RESULT 5
US-08-466-033-106
; Sequence 106, Application US/08466033
; Patent No. 5766840
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungshuk P.
; APPLICANT: Wages, John
; APPLICANT: Young, LaVonne M.
; APPLICANT: Fry, Kirk E.
; APPLICANT: Linnen, Jeffrey M.

RESULT 2
US-09-557-800C-44
; Sequence 44, Application US/09557800C
; Patent No. 6476211
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; APPLICANT: Yeung, George
; TITLE OF INVENTION: Methods and Materials Relating to CD39-Like
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 28110/36457
; CURRENT APPLICATION NUMBER: US/09/557,800C
; CURRENT FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/481,238
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 09/370,265
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/122449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/244444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 44
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: peptide
US-09-557-800C-44
Query Match 12.2%; Score 78; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0018; Mismatches 0; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 93 TRPPRTPLTHE 106
DB 1 TRPPRTPLTHE 14
RESULT 3
US-09-252-991A-27819
; Sequence 27819, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27819
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27819
Query Match 10.5%; Score 67.5; DB 4; Length 96;

TITLE OF INVENTION: Hepatitis G Virus and Molecular
TITLE OF INVENTION: Cloning Thereof
NUMBER OF SEQUENCES: 277
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Ave., Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,033
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/389,886
FILING DATE: 15-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/357,509
FILING DATE: 16-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,729
FILING DATE: 26-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,271
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/285,558
FILING DATE: 03-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/285,543
FILING DATE: 03-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/246,985
FILING DATE: 20-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fadian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 106:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-033-106

Query Match 8.4%; Score 54; DB 1; Length 108;
Best Local Similarity 38.2%; Pred. NO. 33;
Matches 13; Conservative 6; Mismatches 13; Indels 2; Gaps 1;

RESULT 6
US-08-444-733-106
Sequence 106, Application US/08444733
Patent No. 5824507
GENERAL INFORMATION:
APPLICANT: Kim, Jungshuh P.
APPLICANT: Wages, John
APPLICANT: Young, Lavonne M.
APPLICANT: Fry, Kirk E.

APPLICANT: Limen, Jeffrey M.
TITLE OF INVENTION: Hepatitis G Virus and Molecular
TITLE OF INVENTION: Cloning Thereof
NUMBER OF SEQUENCES: 277
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Ave., Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,733
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/389,886
FILING DATE: 15-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/357,509
FILING DATE: 16-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,729
FILING DATE: 26-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,271
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/285,558
FILING DATE: 03-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/285,543
FILING DATE: 03-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/246,985
FILING DATE: 20-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fadian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 106:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-733-106

Query Match 8.4%; Score 54; DB 2; Length 108;
Best Local Similarity 38.2%; Pred. NO. 33;
Matches 13; Conservative 6; Mismatches 13; Indels 2; Gaps 1;

RESULT 7
US-08-464-134-106
Sequence 106, Application US/08464134
Patent No. 5849532
GENERAL INFORMATION:
APPLICANT: Kim, Jungshuh P.
APPLICANT: Wages, John
APPLICANT: Young, Lavonne M.

```

1  APPLICANT: Young, LaVonne M.
2  APPLICANT: Fry, Kirk E.
3  APPLICANT: Linnen, Jeffrey M.
4  TITLE OF INVENTION: Hepatitis G Virus and Molecular
5  TITLE OF INVENTION: Cloning Thereof
6  NUMBER OF SEQUENCES: 277
7  CORRESPONDENCE ADDRESS:
8  ADDRESSEE: Dehlinger & Associates
9  STREET: 350 Cambridge Ave., Suite 250
10 CITY: Palo Alto
11 STATE: CA
12 COUNTRY: USA
13 ZIP: 94306
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: PatentIn Release #1.0, Version #1.25
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/461,361
21 FILING DATE:
22 CLASSIFICATION: 435
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: US 08/389,886
25 FILING DATE: 15-FEB-1995
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: US 08/357,509
28 FILING DATE: 16-DEC-1994
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: US 08/329,729
31 FILING DATE: 26-OCT-1994
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: US 08/344,271
34 FILING DATE: 23-NOV-1994
35 PRIOR APPLICATION DATA:
36 APPLICATION NUMBER: US 08/285,558
37 FILING DATE: 03-AUG-1994
38 PRIOR APPLICATION DATA:
39 APPLICATION NUMBER: US 08/285,543
40 FILING DATE: 03-AUG-1994
41 PRIOR APPLICATION DATA:
42 APPLICATION NUMBER: US 08/246,985
43 FILING DATE: 20-MAY-1994
44 ATTORNEY/AGENT INFORMATION:
45 NAME: Fabian, Gary R.
46 REGISTRATION NUMBER: 33,875
47 REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
48 TELECOMMUNICATION INFORMATION:
49 TELEPHONE: (415) 324-0880
50 TELEFAX: (415) 324-0960
51 INFORMATION FOR SEQ ID NO: 106:
52 SEQUENCE CHARACTERISTICS:
53 LENGTH: 108 amino acids
54 TYPE: amino acid
55 TOPOLOGY: linear
56 MOLECULE TYPE: protein
57 PS-08-461-361-106

```

```

US-06-461-S&L-106
Query Match      8.4%; Score 54; DB 2; Length 108;
Pair Local Similarity 38.2%; Pred No 33;
Matches 13; Conservative 6; Mismatches 13; Indels 2; Gaps 1;

Qy   68 DGHEVVGTMFDAGSTGTRVHVVFQFTRPRRETPT 101
     :::::|::|::|::|::|::|::|::|::|::|
Db   18 EGHAV--GLVSVIHSGGRVTAAETRPWTQVPT 49

```

RESULT 9
US-08-485-910-106
; Sequence 106, Application US/08485910
; Patent No. 5874563
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungsook P
;

APPLICANT: Wages, John
APPLICANT: Young, Lavonne M.
APPLICANT: Fry, Kirk E.
APPLICANT: Linnen, Jeffrey M.
TITLE OF INVENTION: Hepatitis G Virus and Molecular
TITLE OF INVENTION: Cloning thereof
NUMBER OF SEQUENCES: 277
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Ave., Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/485,910
APPLICATION NUMBER: US/08/485,910
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/389,886
FILING DATE: 15-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/357,509
FILING DATE: 16-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,729
FILING DATE: 26-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,271
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/285,558
FILING DATE: 03-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/285,543
FILING DATE: 03-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/246,985
FILING DATE: 20-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 106:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-910-106

Query Match 8.4%; Score 54; DB 2; Length 108;
Best Local Similarity 38.2%; Pred. No. 33;
Matches 13; Conservative 6; Mismatches 13; Indels 2; Gaps 1;

Qy 68 DGHEVYVGIMPDAGSTGTRVHVFOTRPPRETPT 101
Db 18 EGHAV--GMLVSVLHSGGRVTAARTRPWTQVPT 49

RESULT 10
PCT-US95-06266-89
Sequence 89, Application PC/TUS9506266
GENERAL INFORMATION:
APPLICANT:

TITLE OF INVENTION: Detection of Viral Antigens Coded
TITLE OF INVENTION: by Reverse Reading Frames
NUMBER OF SEQUENCES: 157
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06266
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/246,985
FILING DATE: 20-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/285,561
FILING DATE: 03-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,729
FILING DATE: 26-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,271
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/357,509
FILING DATE: 16-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/389,886
FILING DATE: 15-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0202.41
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-06266-89

Query Match 8.4%; Score 54; DB 5; Length 108;
Best Local Similarity 38.2%; Pred. No. 33;
Matches 13; Conservative 6; Mismatches 13; Indels 2; Gaps 1;

Qy 68 DGHEVYVGIMPDAGSTGTRVHVFOTRPPRETPT 101
Db 18 EGHAV--GMLVSVLHSGGRVTAARTRPWTQVPT 49

RESULT 11
US-09-543-681A-6609
Sequence 6609, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILE
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706

```
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6609
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6609

Query Match
  Best Local Similarity 33.3%; Pred. No. 45;
  Matches 17; Conservative 7; Mismatches 19; Indels 8; Gaps 3;

QY 19 LGLCVGVFYVAYIKW---HRATATQAFSITRAAPGARW---GQOAHSP 62
Db 43 LHLCTV-LFTVFKKWKHLKLLASQILSVPLCGDAGSSLVARRHP 92

RESULT 12
US-09-489-039A-8242
; Sequence 8242, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8242
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8242

Query Match
  Best Local Similarity 8.0%; Score 51.5; DB 4; Length 113;
  Matches 25; Conservative 10; Mismatches 35; Indels 27; Gaps 6;

QY 33 KHRATATQAFSITRAAPG---ARWGOAHSLGTAADGHEVFGYIMFDAGSTGTRVHV 89
Db 14 RWRSSAPG---RAGRGSGRRNRPFGDSHP--APSADGIR-----ARSPARRSHA 60

QY 90 FQFT--RPP-----RETPTLTHTFTKAVKPGLS 115
Db 61 GRRARCPPAADGHRRTIRDPPGPAAGCYHAAPPALS 97

RESULT 13
US-09-489-039A-12049
; Sequence 12049, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12049
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12049

Query Match
  Best Local Similarity 8.0%; Score 51; DB 4; Length 101;
  Matches 27; Conservative 7; Mismatches 38; Indels 34; Gaps 4;

; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6609
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6609

Query Match
  Best Local Similarity 33.3%; Pred. No. 45;
  Matches 17; Conservative 7; Mismatches 19; Indels 8; Gaps 3;

QY 19 LGLCVGVFYVAYIKW---HRATATQAFSITRAAPGARW---GQOAHSP 62
Db 43 LHLCTV-LFTVFKKWKHLKLLASQILSVPLCGDAGSSLVARRHP 92

RESULT 14
US-08-959-212-10
; Sequence 10, Application US/08959212
; Patent No. 6060274
; GENERAL INFORMATION:
; APPLICANT: Bjornvad, Made
; APPLICANT: Schulein, Martin
; APPLICANT: Jorgensen, Per
; TITLE OF INVENTION: Extracellular Expression Of Cellulose
; FILE REFERENCE: 4987.200-US
; CURRENT APPLICATION NUMBER: US/08/959,212
; CURRENT FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: 1192/96
; EARLIER FILING DATE: 1996-10-28
; EARLIER APPLICATION NUMBER: 1426/96
; EARLIER FILING DATE: 1996-12-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Humicola insolens
US-08-959-212-10

Query Match
  Best Local Similarity 8.0%; Score 51; DB 3; Length 108;
  Matches 23; Conservative 10; Mismatches 30; Indels 20; Gaps 4;

QY 22 CVGVFIYVAYIKWH-----RATATQAFSITRAAPGARWGOAHSLGTAAD--- 68
Db 23 CVSGYTCVLNDWYVSCQCPQPTTLRTTTTPGATSTTTSAPAA---TSTTPAGCTAERWA 78

QY 69 --GHEVFG-IMPFDAGSTGTRVH 88
Db 79 QCGNGWGGCTTCVAGSTCTKIN 101

RESULT 15
US-09-621-976-4538
; Sequence 4538, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 4538
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 83
; OTHER INFORMATION: Xaa = Ala, Pro
; NAME/KEY: UNSURE
; LOCATION: 15
; OTHER INFORMATION: Xaa = Phe,Leu
```

Tue Jun 8 12:10:42 2004

us-09-905-589a-2_copy_1_120.rat

Page 7

US-09-621-976-4538

```

Query Match 34.78%; Score 50; DB 4; Length 90;
Best Local Similarity 34.0%; Pred. No. 84;
Matches 16; Conservative 5; Mismatches 12; Indels 14; Gaps 2;

Cy 53 ARMGGAASPLGTAD-----GHEVFVGIMFDAGSTGTVYVPCFT 93
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35 ARFGEMSHRGLADPHHTAGAGAGHTTK-----SSRHHVVFST 73

```

Search completed: June 8, 2004, 10:02:46
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model1

Run on: June 8, 2004, 10:06:37 ; Search time 20 Seconds
(without alignments)
1548.684 Million cell updates/sec

Title: US-09-905-589A-2_COPY_135_456
Perfect score: 1659
Sequence: 1 AKQIDFPDMKATPLVKAT.....ALGAIFFHYIDSLNRQSPAS 322

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 167602

Minimum DB seq length: 0
Maximum DB seq length: 322

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_78:1*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	94	5.7	280	2	T28684
2	86	5.2	316	2	D87839
3	86	5.2	316	2	T21451
4	85.5	5.2	244	2	T36357
5	84.5	5.1	246	2	H39999
6	84	5.1	293	2	G72334
7	83	5.0	212	1	S04473
8	82	4.9	212	1	U00706
9	82	4.9	297	2	E36002
10	82	4.9	322	2	T27308
11	81.5	4.9	267	2	AH3234
12	81	4.9	223	2	B37068
13	80.5	4.9	306	1	J01395
14	79.5	4.8	274	1	SRBSD
15	79	4.8	322	2	T24948
16	78.5	4.7	245	2	D72574
17	77.5	4.7	314	2	H64531
18	77	4.6	285	2	AF0146
19	77	4.6	318	2	T50752
20	76.5	4.6	252	2	PC4259
21	75.5	4.6	272	2	S61888
22	75.5	4.6	320	2	E71139
23	75.5	4.6	322	2	S77066
24	75	4.5	250	2	D70787
25	75	4.5	297	2	F33491
26	74.5	4.5	163	2	G95338
27	74.5	4.5	222	2	G72232
28	74.5	4.5	223	2	B71498
29	74.5	4.5	240	2	AC3266

30	74.5	4.5	274	1	D70378	conserved hypothet
31	74.5	4.5	322	2	D97349	glycosyltransferas
32	74	4.5	128	2	T28433	selenophosphate sy
33	73.5	4.4	227	2	S28440	hypothetical prote
34	73.5	4.4	249	2	G75048	proliferating-cell
35	73.5	4.4	284	2	A83688	hypothetical prote
36	73.5	4.4	288	2	D75286	serine proteinase
37	73.5	4.4	301	2	JC2039	phosphoribosylam
38	73.5	4.4	306	2	E72337	hypothetical prote
39	73	4.4	197	2	S28574	ig light chain - A
40	73	4.4	259	2	E95903	probable oxidoredu
41	73	4.4	275	2	T32813	hypothetical prote
42	73	4.4	281	2	D95395	probable Beta lact
43	73	4.4	290	2	D82545	succinyl-CoA synth
44	72.5	4.4	197	2	H90211	conserved hypothet
45	72.5	4.4	249	2	AD2827	short chain dehydr

ALIGNMENTS

RESULT 1
T28684
hypothetical protein - Streptomyces coelicolor
C/Species: Streptomyces coelicolor
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C/Accession: T28684
R:Parkhill, J.; Bentley, S.D.; Barrall, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, January 1999
A/Reference number: Z20512
A/Accession: T28684
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-280 <PAR>
A/Cross-references: EMBL:AL023496; PIDD:CA18901.1
C/Superfamily: hypothetical protein YLR351c

Query Match 5.7%; Score 94; DB 2; Length 280;
Best Local Similarity 31.1%; Pred. No. 1.5;
Matches 37; Conservative 13; Mismatches 53; Indels 16; Gaps 5;
QY 147 GVEGQPAKD---GKEIVSPCLSPFKGEV-----BHAETVYVSGQKAAS-----LHBL 193
Db 136 GNSGWVPEPTAVGKGVYCYDRHPEGRALGLGALVFPASNSRSLGYIMQLBP 195
QY 194 CAARVSEVLQNRVHR--TEEVKHDVPYAFSYYYDLAAG-VGLIDAKGSGSLVGPPEIA 249
Db 196 AAAVNEVFGAINEVGVDELGDNDFFYGRSYVDPEARFVGAVASDKETELVVRDLDNA 254
RESULT 2
D87839
protein F27D4.1 [imported] - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 27-Oct-2003
R:anonymous; The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A/Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A/Reference number: A75000; MUID:19969613; PMID:2851916
A/Note: see website genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_el
A/Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A/Accession: D87839
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-316 <STO>
A/Cross-references: GB:chr_I; PIDD:CA01967.1; PTD:G3876390; GSPDB:GNO0019; CBSP:F27D4.1
C/Genes: C_elegans: F27D4.1
A/Map position: 1
C/Superfamily: electron transfer flavoprotein alpha subunit
Query Match 5.2%; Score 86; DB 2; Length 316;

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C/Accession: T36357

C/R/Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, May 1999

A/Reference number: Z21575

A/Accession: T36357

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-244

A/Cross-references: EMBL:AL049841; PIDN: CAB42784.1; GSPDB: GN00070; SCOEDB: SCE9.38

A/Experimental source: strain A3(2)

C/Genetics:

A/Gene: SCOEDB: SCE9.38

Query Match 5.1%; Score 85.5; DB 2; Length 244;

Best Local Similarity 23.3%; Pred. No. 7.1;

Matches 53; Conservative 23; Mismatches 92; Indels 59; Gaps 9;

QY 67 WITINFLGSLKTPGSSVGMDLGGSTQIAFLPRVEGTLQASPPGY-----LTALRMF 121

DB 20 WVTAGTASTENTASAESTS---GAWRTGVA-----AGTASATAPGTDVRSGGVAFRRS 70

QY 122 NRTYKLYSVSYLGLGLMRLAIL-----GGVEGQPAKDGKELVSPCLSPS----- 167

DB 71 RYTRSRVVAALSLGLVGA--LALACTDGGSDGGSGEGSGDKSSAAPGNASATDPGG 129

QY 168 -----FKGEV-----EHAETVTVSGQKAA--ASLHELCAARSEVLQNRVHR 208

DB 130 DAGGSFSAAGAELESGSLATTDGQAVALMTGDKAALFATGGTVCSGTTTSGTRTIR 189

QY 209 -----TEEVKHVDYFAFSYYYDLAAGVGLIDAEGKGSIVVG 244

DB 190 LKCADGSADRTCKVGAVGATSLTVANEGALGKETVYTRSEGGSLPPG 236

RESULT 5

H39999

hypothetical protein 8 - Chlamydomophila psittaci

C/Species: Chlamydomophila psittaci, Chlamydia psittaci

C/Date: 08-Nov-1991 #sequence_revision 05-May-2000 #text_change 09-Dec-2002

C/Accession: H39999; S18149; S18141

R/Lusher, M.E.; Gregory, J.; Storey, C.C.; Richmond, S.J.

submitted to the Protein Sequence Database, October 1991

A/Reference number: A39999

A/Accession: H39999

A/Status: not compared with conceptual translation

A/Molecule type: DNA

A/Residues: 1-246

R/Lusher, M.E.; Gregory, J.; Storey, C.C.; Richmond, S.J.

submitted to the EMBL Data Library, October 1991

A/Description: Analysis of the complete nucleotide sequence of the plasmid pCpal isolate

A/Reference number: S18141

A/Accession: S18149

A/Molecule type: DNA

A/Residues: 1-99

A/Cross-references: EMBL:X62475; NID:g40608; PIDN:CAA44340.1; PID:g581000

A/Accession: S18141

A/Molecule type: DNA

A/Residues: 100-246

A/Cross-references: EMBL:X62475; NID:g40608; PIDN:CAA44332.1; PID:g40609

C/Genetics:

A/Start codon: GTG

C/Superfamily: Virulence pGP6-D-related protein

Query Match 5.1%; Score 84.5; DB 2; Length 246;

Best Local Similarity 20.7%; Pred. No. 8.8;

Matches 44; Conservative 40; Mismatches 84; Indels 45; Gaps 10;

[illegible]

```

Db      114 FRT---KSAAYNALGYEL-----FTSLDPKTKTSIFOSIP--YNTAYILASRK----- 157
Qy      130 YSYLGLGMSARLAILGVGVGPAPKDGKELVSPCLSPSPFGMEWHAETVTVSGQRAAS 189
Db      158 -----GSVXDKLKVLKIKISGLSNALADIVLNKFLPPLKSSQTERC-VDFEEKKEVSEK 210
Qy      190 LHEL-----CAARSEVTLQNRVHRT--EYVGHND 216
Db      211 LIDILKIVSGGLEISEYKNKLLHOLFETKLKVD 243

```

RESULT 6

```

G72334 [acyl-carrier-protein] S-malonyltransferase (EC 2.3.1.39) TM0798 [similarity] - Thermoc
C:Species: Thermocoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: G72334
R:Neison, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.V.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: G72334
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-293 <AB>
A:Cross-references: GB:AA001747; GB:AA000512; NID:54981304; PIDD:AA035680.1; PID:g498132
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0798
C:Superfamily: [acyl-carrier-protein] S-malonyltransferase; [acyl-carrier-protein] S-mal
C:Keywords: acyltransferase; coenzyme A
P:3781/Dominant [acyl-carrier-protein] S-malonyltransferase homology <AMT>
P:87/Active site: Ser (covalent substrate-binding) #status predicted
P:193/Active site: His #status predicted

```

```

Query Match      5.1%; Score 84; DB 2; Length 293;
Best Local Similarity 24.2%; Pred. No. 12;
Matches 59; Conservative 32; Mismatches 67; Indels 86; Gaps 13;

Qy      25 LIPGKAKQ-----KLLQKVEYFKASPLVGDCCSINMGDEGV-----SAMI 68
Db      5 VFPGGSGSYGMAKDPFVYESKKEIFERSKVLGPDITENNGDEFTLKTENAPSIYI 64
Qy      69 TINFITGSLKTPGGSSVGMGLDGGSGTIAFLPRVREGTLQASPGVYLTALMFRRTKLY 128
Db      65 T-----SYIAFLEI-----EKRGILPDV---VAHSHLGEYTLALAVAG-VDFE 103
Qy      129 SYSYV-----GLGMSARLAILGVGVGPAPKDGKELVSPCLSPSPFGMEWHAETVTVSGQRAAS 172
Db      104 TGLYLVRRKGEYMSKALPEKGTMAA-----VGLNIETIEEVN-----SIEGVYIA 151
Qy      173 ---EHAETTVYSG-----QKRAASLHELCAARSEVTLQNRVHRT-----BEVK 213
Db      152 NYNSHDQVV--ISGLKSEVERKAMETLKEKARAVVELVWVSPFHTPLFLEVRKMKEEVE 209
Qy      214 HVDF 217
Db      210 KYDF 213

```

RESULT 7

```

S04473 nitrille hydratase (EC 4.2.1.84) beta chain [validated] - Rhodococcus sp.
C:Species: Rhodococcus sp.
C>Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 27-Oct-2003
C:Accession: S04473; S02071; S15072; S54106; S62096
R:Ikehata, O.; Nishiyama, M.; Horiouchi, S.; Beppu, T.
Eur. J. Biochem. 181, 563-570, 1999
A>Title: Primary structure of nitrille hydratase deduced from the nucleotide sequence of
A:Reference number: S04471; MUID:89276338; PMID:2659343

```

```

A:Accession: S04473
A:Molecule type: DNA
A:Residues: 1-212 <IK>
A:Cross-references: EMBL:X14668; NID:g46429; PIDD:CA432798.1; PID:g46431
A>Note: part of this sequence, including the amino end of the mature protein, was confir
R:Endo, T.; Watanabe, I.
FEBS Lett. 243, 61-64, 1989
A>Title: Nitrille hydratase of Rhodococcus sp. N-774. Purification and amino acid sequenc
A:Reference number: S02070; MUID:89153549; PMID:2920826
A:Accession: S02071
A:Molecule type: protein
A:Residues: 1-19 <END>
R:Hashimoto, Y.; Nishiyama, M.; Ikehata, O.; Horiouchi, S.; Beppu, T.
Biochim. Biophys. Acta 1088, 225-233, 1991
A>Title: Cloning and characterization of an amidase gene from Rhodococcus species N-774
A:Reference number: S15070; MUID:91159474; PMID:2001397
A:Accession: S15072
A:Molecule type: DNA
A:Residues: 1-12/179-212 <HA>
A:Cross-references: EMBL:X54074
A:Experimental source: strain N-774
R:Bigey, F.; Chebrou, H.; Arnaut, A.; Galay, P.
submitted to the EMBL Data Library, March 1995
A:Description: Cloning, sequencing of the modified nitrille hydratase gene from mutant st
A:Reference number: S54104
A:Accession: S54106
A:Molecule type: DNA
A:Residues: 1-39, 'V', 41-212 <BIG>
A:Cross-references: EMBL:Z48769; NID:g769823; PIDD:CA48686.1; PID:g769826
R:Rhang, W.; Schneider, G.; Lindqvist, Y.
submitted to the Brookhaven Protein Data Bank, April 1997
A:Reference number: A73039; PDB:1AHO
A:Contents: annotation, X-ray crystallography, 2.65 angstroms
R:Nagashima, S.; Nakasako, M.; Dohmae, N.; Tsujimura, M.; Takio, K.; Odaka, M.; Yoshida,
Nat. Struct. Biol. 5, 347-351, 1998
A>Title: Novel non-heme iron center of nitrille hydratase with a claw setting of oxygen e
A:Reference number: A58907; MUID:98246406; PMID:9586994
A:Contents: annotation, X-ray crystallography, 1.7 angstroms; mass spectroscopic identifi
A:Function:
C:Description: catalyzes the reaction of one molecule of water and an aliphatic nitrile
C:Superfamily: nitrille hydratase; beta subunit
C:Keywords: carbon-oxygen lyase; hydro-lyase; iron; metalloprotein
P:1-212/Product: nitrille hydratase beta chain #status experimental <MAT>
P:55,111/Binding site: nitrosyl iron (Arg) (shared with alpha chain) #status experimente

```

```

Query Match      5.0%; Score 83; DB 1; Length 212;
Best Local Similarity 29.0%; Pred. No. 9.7;
Matches 38; Conservative 17; Mismatches 46; Indels 30; Gaps 8;

```

```

Qy      145 LGVNEGPAKDGK--ELVSPCLSPSPFGMEWHAETVTVSGQRAASLHELCAARSEVTL 202
Db      7 LAGVQG-----FGKVPRTVADIGPTFAHWEHLPSLMAAG-----VAELGAFVDEYV 55
Qy      203 QNRVHRTBEYKVDYFAFSYYD-LAAGVGLIDAEKG-----GSLVGDPEIARVYC 254
Db      56 -RYVERMEPRH---YMTPIYERYVIGVATLMVERGILTDQLSLAGGPPLS----- 106
Qy      255 RILETQPOSSP 265
Db      107 RPSESEGRAP 117

```

RESULT 8

```

JN0706 nitrille hydratase (EC 4.2.1.84) beta chain - Rhodococcus erythropolis
C:Species: Rhodococcus erythropolis
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 27-Oct-2003
C:Accession: JN0706
R:Duran, R.; Nishiyama, M.; Horiouchi, S.; Beppu, T.
Biosci. Biotechnol. Biochem. 57, 1323-1328, 1993
A>Title: Characterization of nitrille hydratase genes cloned by DNA screening from Rhodoc
A:Reference number: P06707; MUID:93379358; PMID:7764017
A:Accession: JN0706

```

T27308
hypothetical protein Y69E1A.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T27308
R:Lennard, N.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z20342
A:Accession: T27308
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-322 <MIL>
A:Cross-references: EMBL:AL034365; PIDN:CAA22256.1; CESP:Y69E1A.1
C:Genetics:
A:Gene: CESP:Y69E1A.1
A:Introns: 71/1; 291/3
C:Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.1

	Query Match	4.9%;	Score 82;	DB 2;	Length 322;
	Best Local Similarity	25.6%;	Pred. No. 21;		
	Matches	54;	Conservative 29;	Mismatches 78;	Indels 50; Gaps 10;

QY 38 VKEVFK-ASFLVGDVDCVIMNGTDEGVSAWITINFLTGSLKTPGQSSV-----GM 87
DB 119 VVEIFGWSILLGAGIAGIIG-----YVLSPIFGIFGRGGSAILATLAIPAFGA 170
QY 88 LDIGG--GST-----QIAFLPRVEGT-----LQASPPGYLTALRMFNRTYKLYSY 130
DB 171 YQLNEDGSGTGRFQLLVLLVQGLLMGHSISYTYLSAQLPGVTPLVI-----AFAY 224
QY 131 SYLGLGLMSARLAILGGVEGQPAKDGKEIVSPCLSPFKGEWEHAEVTVRVSGQKAASL 190
DB 225 PLIAGQVGTAFTPLIGGAVG--AAFQTQLVLGLVSGLSFSYLLLSALY-----SAASG 276
QY 191 HELCAARVSEVLQNRVHRTVEVKHVDYAFS 221
DB 277 ALLQVAPKNTAQNRH-MYQILLVSSFLFS 306

RESULT 11
AH3234
hypothetical protein seub [imported] - Agrobacterium tumefaciens (strain C58, Dupont) pl
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AH3234
R:Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
i; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AH3234
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-267 <KUR>
A:Cross-references: GB:AB008690; PIDN:AAL46294.1; PID:gl17744077; GSPDB:GN00189
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: seub
A:Genome: plasmid

	Query Match	4.9%;	Score 81.5;	DB 2;	Length 267;
	Best Local Similarity	23.6%;	Pred. No. 18;		
	Matches	65;	Conservative 28;	Mismatches 87;	Indels 95; Gaps 14;

QY 42 FXASPLFVGDVDCVIMNGTDEGVSAWITINFLTGSLKTPGSSVGMMLD-----LGGGSGTQ 96
DB 10 FPAITPAADTHLLV-----QGVSK-----SFSRPGDSEVRALDHVDLVVRDGS-- 53
QY 97 IAFILPRVEGTLOASPPGYLTALRMFNRTYKLYSYLGLMSARLAILGGVE-----GQP 152

Db 54 -----VCTTGAAGCGGKSTLLR-----IVAGLEPQFGGTV 83
 QY 153 AKDGKELVSPCLSPSPFKGEMHAEVTVKVSOKAAASLHELCAARVSEVLQNRVHRTVEV 212
 Db 84 LLGGKPLGPGDLGRIIVQ-DHRLVPMWTVEANLAFSLHRLPKAEGRRV-----TEKL 136
 QY 213 KHVDFYAF--SYIIDLAAGVGLIDAEKGGSLVGDFFELIAKYVCTETQPO---SSPF 266
 Db 137 KLVGLEGRSPYHQLSGMAQRYA-----IARALAHQPELLLIDEPF 179
 QY 267 SCMD-LTVYSLLTLEFGFPRSKVLTETKIDNVER 300
 Db 180 GADLALTRLOM-----QDEVLRI-RHTDNLTT 205

RESULT 12

B97068

cobalamin biosynthesis protein CblM [Imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001

C:Accession: B97068

R:Noiling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B97068

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-223 <KUR>

A:Cross-references: GBA:001437; PIDN:AAK79333.1; PID:G15024300; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

A:Genetics:

A:Gene: CAC1365

C:Superfamily: cobalamin biosynthesis protein M

Query Match

Best Local Similarity 23.4%; Score 81; DB 2; Length 223;

Matches 43; Conservative 28; Mismatches 71; Indels 42; Gaps 7;

QY 17 LKATAGRLLPBKKAKLQKYEVPKASPF---VGDGCVSINMGIDEGVSANITTF 72
 Db 16 LVAAASLFFVPG--VREIKRSEKENDLYKPFISMVGVPVISCW-----IYVP- 63
 QY 73 LKSLKTPGSSVGMGLDGGSTQIAFLPRVEGTQAQSPGYLALRFNFTYKLYSY 132
 Db 64 VTGSGCHPGCTPLAAILIGPLATSV-----ITIALFPQALFLGHGCI 106
 QY 133 LGLMSARLAILGVEGP---AKDGKELVSPCLSPSPFKGEMHAEVTVKVSOKAA 187
 Db 107 TTIGANNISMGIAISGYAFKLPFKFGSSVWLAAGVAGVGDW---VTYVMSALELA 162
 QY 188 ASLH 191
 Db 163 VSLH 166

RESULT 13

J01395

phosphoribosylaminoimidazoleleucinecarboxamide synthase (EC 6.3.2.6) - yeast (Saccharomy

N:Alternate names: protein YAR015W; SAICAR synthetase

C:Species: Saccharomyces cerevisiae

C:Date: 17-Jul-1992 #sequence_revision 03-Nov-1995 #text_change 05-Nov-1999

C:Accession: S20122; S40905; S25679; J01395

R:Davies, C.V.; Hutchison III, C.A.

Nucleic Acids Res. 19, 5731-5738, 1991

A:Title: A directed DNA sequencing strategy based upon Tn3 transposon mutagenesis: appli

A:Reference number: S20121; MUID:92051323; PMID:1658741

A:Accession: S20122

A:Molecule type: DNA

A:Residues: 1-306 <DAV>

A:Cross-references: EMBL:M67445; NID:g170997; PIDN:AAA34398.1; PID:g170999

R:Clark, M.W.; Keng, T.; Storms, R.K.; Zhong, W.; Fortin, N.; Zeng, B.; Delaney, S.; Que
 submitted to the EMBL Data Library, November 1993

A:Description: Sequencing of chromosome I of Saccharomyces cerevisiae: Analysis of the

A:Reference number: S40905

A:Accession: S40905

A:Molecule type: DNA

A:Residues: 1-306 <CLA>

A:Cross-references: EMBL:122015; NID:g1339990; PIDN:AA04963.1; PID:g349755; GSPDB:GN00

R:Schweitzer, B.; Phillips, P.

Mol. Gen. Genet. 234, 164-167, 1992

A:Title: NPK1, a nonessential protein kinase gene in Saccharomyces cerevisiae with simi

A:Reference number: S23580; MUID:92357012; PMID:1495480

A:Accession: S23580

A:Molecule type: DNA

A:Residues: 285-306 <SCH>

A:Cross-references: EMBL:X60549; NID:g298024; PIDN:CAA43043.1; PID:g4043

R:Myasnikov, A.N.; Sasnauskas, K.V.; Janulaitis, A.A.; Smirnov, M.N.

Gene 109, 143-147, 1991

A:Title: The Saccharomyces cerevisiae ADE1 gene: structure, overexpression and possible

A:Reference number: J01395; MUID:92097943; PMID:1756975

A:Accession: J01395

A:Molecule type: DNA

A:Residues: 1-184, 'G', 186-306 <MTA>

A:Cross-references: GB:M61209; NID:g170995; PIDN:AAA34396.1; PID:g170996

A:Experimental source: strain FH4C

A:Note: the authors translated the codon TAC for residue 98 as Asp

C:Genetics:

A:Gene: SGD:ADE1; MIPS:YAR015W

A:Cross-references: SGD:S0000070; MIPS:YAR015W

A:Map position: 1R

C:Function:

A:Description: catalyzes formation of N-succinyl-5-aminoimidazole-4-carboxamide ribotide

A:Pathway: purine biosynthesis

C:Superfamily: phosphoribosylaminoimidazoleleucinecarboxamide synthase

C:Keywords: ligase; purine nucleotide biosynthesis

Query Match

Best Local Similarity 21.8%; Score 80.5; DB 1; Length 306;

Matches 58; Conservative 40; Mismatches 91; Indels 77; Gaps 16;

QY 37 KXKEVFPASP---FLVGDGCVS---INNGT---DEGV-----SAWITINFLGSLKTPG 81
 Db 19 KVRDIIEVDAGTLFLPATDRISAIVYIMENSIPKILITKSEFF--FKLSNDVKN-- 74
 QY 82 GSSVGMGLDGGSTQIAFLPRVEGTQAQSPGYLALRFNFTYKLYSY 135
 Db 75 ----HLVDIAPGKTITFDYLP---AKLSBPKYKTLG--DRSLVHKKILPLEVTVRG 123
 QY 136 ----GLMSARLAILGCVBG---QPADGKELVSPCLSPSPFKGE-WHAETVTVKVSOKAA 187
 Db 124 YITGSAKVEYKGTGYHAKQPGQLKESQFPPEPITPTKAEQGHDE--NISPAQAA 180
 QY 188 ASLHELCAARVSEVLQNRVHRTVEVHVDFYAFSYYYDIAAGVGLIDA-----EKGS 240
 Db 181 ELVGEDLSRRVAFIA-----VLYSKCKRYAKKXGILLIDTFEFGIDKTNF 228
 QY 241 LVVGDFFIAKYVCTLETOPQSSPF 266
 Db 229 IIVVD-EVLT-----PDSSRF 243

RESULT 14

SUBSD

subtilisin (EC 3.4.21.62) DY - Bacillus subtilis (strain DY)

N:Alternate names: alkaline serine proteinase

C:Species: Bacillus subtilis

A:Variety: strain DY

C:Date: 20-Sep-1984 #sequence_revision 20-Sep-1984 #text_change 02-Jul-1998

C:Accession: A00969; S02492

R:Nedkov, P.; Oberthur, W.; Braunitzer, G.

Biochem. Hoppe-Seyler 366, 421-430, 1985

A:Title: Determination of the complete amino-acid sequence of subtilisin DY and its com

A:Reference number: A00969; MUID:85279896; PMID:3927935

Search completed: June 8, 2004, 10:10:38
Job time : 22 secs

```

RESULT 15
T24948
hypothetical protein T16A9.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T24948
R:McMurray, A.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z19959
A:Accession: T24948
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-322 <W11>
A:Cross-references: EMBL:Z77135; PIDN:CA800878.1; GSPDB:GN00023; CESP:T16A9.5
A:Experimental source: clone T16A9
C:Genetics:
A:Gene: CESP:T16A9.5
A:Map position: 5
A:Introns: 71/1, 291/3
C:Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.1

Query Match 4.8%; Score 79; DB 2; Length 322;
Best Local Similarity 25.1%; Pred. No. 38;
Matches 53; Conservative 30; Mismatches 78; Indels 50; Gaps 10;

QY 38 VKPEVK-ASPLVGDGDCVSMNGTDEGVSAWITNFIQTGLKTPGGSSV-----GM 87

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 10:02:51 ; Search time 18 Seconds

(Without alignments)
931.477 Million cell updates/sec

Title: US-09-905-589a-2_COPY_135_456
Perfect score: 1659
Sequence: 1 AKODIPDFPKATPLVLKAT.....ALGAFHYIDSLNRKSPAS 322

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 76390

Minimum DB seq length: 0
Maximum DB seq length: 322

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84.5	5.1	246	1	GP6D_CHLPS
2	83	5.0	212	1	NPAB_RHOER
3	82.5	5.0	192	1	FM11_HUMAN
4	81.5	4.9	312	1	GBLP_LEICR
5	81.5	4.9	312	1	GBLP_LEICR
6	79.5	4.8	274	1	SUBD_BACLI
7	78	4.7	287	1	YQ73_VIBPA
8	77	4.6	315	1	ENGC_PORGI
9	77	4.6	318	1	BCHC_RHOSE
10	75.5	4.6	286	1	CLN8_HUMAN
11	75.5	4.6	322	1	V712_SIN3
12	75	4.5	250	1	YMI9_MYCTU
13	74.5	4.5	274	1	PPNK_AOUAE
14	73.5	4.4	227	1	YFVA_MERTF
15	73.5	4.4	249	1	PCNA_PYRAB
16	73.5	4.4	301	1	PUR7_PICIA
17	73.5	4.4	305	1	SYGA_STRPY
18	73.5	4.4	306	1	PUR7_YEAST
19	73.5	4.4	308	1	SYGA_STRPY
20	72.5	4.4	303	1	PUR7_PICAN
21	72.5	4.4	303	1	SUBH_RHIME
22	72.5	4.4	305	1	SYGA_STRPY
23	72	4.3	254	1	ADH_DBOLP
24	72	4.3	276	1	YGF8_ECOLI
25	72	4.3	316	1	LDH_BORBU
26	71.5	4.3	293	1	ALF1_PORGI
27	71	4.3	163	1	DEF3_SHEON
28	70.5	4.2	292	1	YG38_PYRPU
29	70.5	4.2	125	1	KHUI_HUMAN
30	70.5	4.2	275	1	KHUI_HUMAN
31	70.5	4.2	284	1	TPM2_CHICK
32	70	4.2	183	1	OLEC_BRANA
33	70	4.2	258	1	FAB1_ANASP

ALIGNMENTS

RESULT 1	GP6D_CHLPS	STANDARD;	PRT;	246 AA.
ID	GP6D_CHLPS			
AC	Q46264; Q46256;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Virulence plasmid protein GP6-D.			
OS	Chlamydia psittaci (Chlamydia psittaci).			
OG	Plasmid pCPA1.			
OC	Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX	NCBI_TaxID=83554;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=N352;			
RX	MEDLINE=97346036; PubMed=9202459;			
RA	Thomas N.S., Lusher M., Storey C.C., Clarke I.N.;			
RT	"Plasmid diversity in Chlamydia."			
RL	Microbiology 143:1847-1854(1997).			
CC	-1- SIMILARITY: BELONGS TO THE UPF0137 (GP6-D) FAMILY.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL collabora-			
CC	tion the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb.ch/announce/			
CC	or send an email to license@isb.ch).			
CC	-----			
DR	EMBL; X62475; CAA44340.1; -			
DR	EMBL; X62475; CAA44332.1; -			
DR	PIR; H39999; H39999			
DR	InterPro; IPR003350; UPF0137.			
DR	PIfam; PF03677; UPF0137, 1.			
KW	Plasmid.			
SQ	SEQUENCE 246 AA; 28110 MW; 78366B3FA4152DFP CRC64;			
Query Match	5.1%; Score 84.5; DB 1; Length 246;			
Best local similarity	20.7%; Pred. No. 5.7;			
Matches	44; Conservative 40; Mismatches 84; Indels 45; Gaps 10;			
QY	18 KATAGIRLIPBEKAQKLQKVEFKASPPFVGDVSI-----MNGTDEGVSAWITIN 71			
DB	62 RGLASIKILTQD-----IKSIQKQVLLIGEKIVKREILKNMNSPDTTFSWIMLV 113			
QY	72 PLTSGIKTPGGSVGMULDGGSGVQIARLP--RVESGTQASPPVYTLTMRNRTYKYS 129			
DB	114 FRT--KSAVNAAGYIEL-----FSLPDKNTKSLFQSTP--YKTAVALASKR----- 157			
QY	130 YSYLGLMSARLILGVEGQPAKDGKELVSPCLSPFKEMWHAETVYVSGQKAAS 189			
DB	158 -----GSVKOKLKVLTGKISGNALADIVANKFLPPLKSSQTRC--VDPEKKKEVSEK 210			
QY	190 LHEL-----CAARVEVQNRVHRT--EEVKHYD 216			
DB	211 LIDILKIVSGGLESEYKKNLHLQLEFKTLKVD 243			

```

RESULT 2
NHAB_RHOER
ID NHAB_RHOER STANDARD; PRT; 212 AA.
AC P13449; Q59789;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nitrile hydratase subunit beta (EC 4.2.1.84) (Nitrilase) (NHase).
GN NTHB OR NHA2.
OS Rhodococcus erythropolis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=N-774;
RX MEDLINE=89276338; PubMed=2659343;
RA Ikemata O., Nishiyama M., Horinouchi S., Beppu T.;
RT "Primary structure of nitrile hydratase deduced from the nucleotide
RT sequence of a Rhodococcus species and its expression in Escherichia
RT coli.";
RL Eur. J. Biochem. 181:563-570(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=N-774;
RX MEDLINE=91159474; PubMed=2001397;
RA Hashimoto Y., Nishiyama M., Ikemata O., Horinouchi S., Beppu T.;
RT "Cloning and characterization of an amidase gene from Rhodococcus
RT species N-774 and its expression in Escherichia coli.";
RL Biochim. Biophys. Acta 1088:225-233(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ACV2;
RA Bigey F., Chabrou H., Arnaud A., Galzy P.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=N-771;
RA Nojiri M., Yoshida M., Odaka M., Matsushita Y., Tsujimura M.,
RA Yoshida T., Takio K., Endo I.;
RT "Functional expression of nitrile hydratase in Escherichia coli.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-19.
RC STRAIN=N-774;
RX MEDLINE=89153549; PubMed=2920826;
RA Endo T., Watanabe I.;
RT "Nitrile hydratase of Rhodococcus sp. N-774. Purification and amino
RT acid sequences.";
RL FEBS Lett. 243:61-64(1989).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.65 ANGSTROMS).
RC STRAIN=Brevibacterium sp. R312;
RX MEDLINE=97341231; PubMed=9195885;
RA Huang W., Jia J., Cummings J., Nelson M., Schneider G., Lindqvist Y.;
RT "Crystal structure of nitrile hydratase reveals a novel iron centre
RT in a novel fold.";
RL Structure 5:691-699(1997).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS), AND MASS SPECTROMETRY.
RX MEDLINE=98246406; PubMed=9586994;
RA Nagashima S., Nakasako M., Dohmae N., Tsujimura M., Takio K.,
RA Odaka M., Yoshida M., Kamiya N., Endo I.;
RT "Novel non-heme iron center of nitrile hydratase with a claw setting
RT of oxygen atoms.";
RL Nat. Struct. Biol. 5:347-351(1998).
CC -1- FUNCTION: NHASE CATALYZES THE HYDRATION OF VARIOUS NITRILE
CC COMPOUNDS TO THE CORRESPONDING AMIDES. INDUSTRIAL PRODUCTION OF
CC ACRYLAMIDE IS NOW BEING DEVELOPED USING SOME OF THESE ENZYMES.
CC -1- CATALYTIC ACTIVITY: An aliphatic amide = a nitrile + H(2)O.
CC -1- SUBUNIT: Heterodimer of an alpha and a beta chain.

```

```

-1- SIMILARITY: Belongs to the nitrile hydratase subunit beta family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X54074; CAA38011.1; -
DR EMBL; X14668; CAA32798.1; -
DR EMBL; Z48769; CAA86686.1; -
DR EMBL; AB016078; BAA36598.1; -
DR PDB; 1AHJ; 08-APR-98.
DR PDB; 2AHJ; 16-FEB-99.
DR InterPro; IPR008990; E transp acc.
DR InterPro; IPR003168; NHase_beta.
DR Pfam; PF02211; NHase_beta_1.
DR PRSFP; PRSFP001427; NHase_beta; 1.
KW Lyase; 3D-structure.
FT VARIANT 40 5 M -> V (IN STRAIN ACV2).
FT TURN 4 5
FT TURN 8 9
FT TURN 20 21
FT TURN 31 34
FT TURN 35 45
FT TURN 46 46
FT TURN 52 60
FT TURN 61 61
FT TURN 64 69
FT HELIX 72 86
FT TURN 87 88
FT TURN 92 99
FT TURN 100 100
FT TURN 125 126
FT STRAND 128 131
FT HELIX 144 146
FT TURN 147 148
FT STRAND 150 156
FT STRAND 161 161
FT HELIX 164 167
FT TURN 168 170
FT STRAND 177 177
FT STRAND 179 185
FT HELIX 186 190
FT TURN 191 192
FT STRAND 198 204
FT HELIX 205 207
FT STRAND 208 210
SQ SEQUENCE 212 AA; 23487 MW; A0401CAAFC1C2CBE CRC64;
Query Match 5.0%; Score 83; DS 1; Length 212;
Best Local Similarity 29.0%; Pred. No. 6.4; Indels 30; Gaps 8;
Matches 38; Conservative 17; Mismatches 46;
Qy 145 LGVEGQAPKDGK--ELVSPCLSPSPFKGEWEHAETVYRSGQKAAASLHELCAARVSEVL 202
Db 7 LAGVQG---FGKVPHTVNADIGTPTFAEWEHLPSYLMFAG-----VAELGAFSVDEV- 55
Qy 203 QNEVHRTVEVKHVDYFAFSVYTD-LAAGVGLIDAEGK-----GSLVVGDEIARVYC 254
Db 56 -RVVVERMEPRH---YMMTPYERYVIGVATLMVKEGILTQDELESLAGPPPLS----- 106
Qy 255 RTLETQPSQSP 265
Db 107 RPSESEGRGAP 117

```

```

RESULT 3
RM11 HUMAN
ID RM11 HUMAN STANDARD; PRT; 192 AA.
AC Q9Y3B7; Q96Q73;

```

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE 60S ribosomal protein L11, mitochondrial precursor (L11mt) (CGI-113).

GN MRP11.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RP MEDLINE=20272150; PubMed=10810093;

RA Lai C.-H., Chou C.-Y., Chang L.-Y., Liu C.-S., Lin W.-C.;

RT "Identification of novel human genes evolutionarily conserved in

RT Caenorhabditis elegans by comparative proteomics.";

RL Genome Res. 10:703-713 (2000).

RL [2]

RP SEQUENCE FROM N.A.

RP MEDLINE=21293042; PubMed=11279069;

RA Suzuki T., Terasaki M., Takemoto-Hori C., Hanada T., Ueda T., Wada A.,

RA Watanabe K.;

RT "Structural compensation for the deficit of rRNA with proteins in the

RT mammalian mitochondrial ribosome. Systematic analysis of protein

RT components of the large ribosomal subunit from mammalian

RT mitochondria.";

RL J. Biol. Chem. 276:21724-21736 (2001).

RL [3]

RP SEQUENCE FROM N.A.

RP TISSUE=Lymph;

RP MEDLINE=22388257; PubMed=12477932;

RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins P.S., Wagner L., Shemen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buerger K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stepieton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.U., Ustin T.B., Toshiyuki S., Carrinci P., Prange C.,

RA Raha S.S., Locellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

RA Bosak S.A., Mesman P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Halik S.W.,

RA Villalón D.K., Murry D.M., Sodergren E.U., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,

RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length

RT human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RL [4]

RP SEQUENCE FROM N.A.

RP MEDLINE=21429115; PubMed=11543634;

RA Kameuchi N., Suzuki T., Uechi T., Magoori M., Kuniba M., Higa S.,

RA Watanabe K., Tanaka T.;

RT "The human mitochondrial ribosomal protein genes: mapping of 54 genes

RT to the chromosomes and implications for human disorders.";

RL Genomics 77:65-70 (2001).

RL -1- SUBCELLULAR LOCATION: Mitochondrial.

CC -1- SIMILARITY: Belongs to the L11P family of ribosomal proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL, AF151871; AAD34108.1; -

DR EMBL, AB048638; BAB40843.1; -

DR EMBL, BC005002; AAH05002.1; -

DR EMBL, AB051338; BAB54928.1; -

DR HESP, P29395; IMMS.

DR Genew; HGNC:14042; MRP11.

DR InterPro; IPR000911; Ribosomal_L11.

DR InterPro; IPR006519; Ribosomal_L11bac.

DR Pfam; PF002398; Ribosomal_L11; 1.

DR Pfam; PF03946; Ribosomal_L11N; 1.

DR Pfam; PF001367; Ribosomal_L11; 1.

DR SMART; SM00649; RLI1; 1.

DR TIGRFAMs; TIGR01632; L11_bact; 1. FALSE NEG.

DR PROSITE; PS00359; RIBOSOMAL_L11.

DR Ribosomal protein; Mitochondrion; Transl. peptid.

FT TRANSIT 1 ?

FT CHAIN ? 192

FT SEQUENCE 192 AA; 20683 MW; 92506A2004AB278 CRC64;

Query Match 5.0%; Score 82.5; DB 1; Length 192;

Best Local Similarity 22.7%; Pred. No. 6.2;

Matches 42; Conservative 28; Mismatches 76; Indels 39; Gaps 8;

QY 131 SYLGGLMSRLILGVGEQPKDQKELVSPCLSPFKGEMHAEVTVRVSQKAAASL 190

DB 2 SKLGRARGLRKEVGVIRAIYRAGLAMPPLGP-----VLGQR-GVSI 46

QY 191 HELCA--ARVSEVLQN-----RVHTEEVKVDFAFSGYYDLAAGVLIDAEK 238

DB 47 NQCKEENERTKDIKIGIPPTKILVYKPRTEPIK-IGQPTSYFLKAAAGT-----EK 100

QY 239 ----GSLVGDPEIATKYCRITLPTOPQSPSCMDLTVYSLIQFGPPRSKYLKTR 294

DB 101 ARQTKVEKVAQ--LVTLKHVEIRAIKQDEAFALQVPLSSVRSIIIGARSIGIRVXD 158

QY 295 IDNVE 299

DB 159 LSSEB 163

RESULT 4

GBLP LEICH STANDARD; PRT; 312 AA.

ID 027434;

AC 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Granule nucleotide-binding protein beta subunit-like protein (Antigen

DE LACK) (LIP36) (P36L1).

OS Leishmania chagasi, and

OS Leishmania infantum.

OC Eukaryota; Eumetazoa; Kinetoplastida; Trypanosomatidae; Leishmania.

OX NCBI_TaxID=44271, 5671;

OX [1]

RP SEQUENCE FROM N.A.

RP SPECIES=L.chagasi;

RP MEDLINE=95242101; PubMed=7725103;

RA Mougneau E., Altare F., Makil A.E., Zheng S., Coppola T., Wang Z.E.,

RA Waldmann R., Locksley R.M., Glachanhaus N.;

RT "Expression cloning of a protective Leishmania antigen.";

RL Science 268:563-566 (1995).

RL [2]

RP SEQUENCE FROM N.A.

RP SPECIES=L.infantum; STRAIN=PB75;

RA Gonzalez-Aleguino J., Larraga V.;

RT Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.

RL -1- DEVELOPMENTAL STAGE: EXPRESSED IN BOTH STAGES OF THE PARASITE LIFE

CC CYCLE.

CC -1- SIMILARITY: Contains 7 WD repeats.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

```
CC ENBL; U27568; AAA57576.1; -
DR ENBL; U49695; AAA51208.1; -
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINERPT.
DR PRODOM; PD000018; WD40; 4.
DR SMART; SM00320; WD40; 6.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00882; WD_REPEATS_2; 5.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat; Antigen.
FT REPEAT 9 42 WD 1.
FT REPEAT 63 93 WD 2.
FT REPEAT 105 135 WD 3.
FT REPEAT 148 180 WD 4.
FT REPEAT 192 222 WD 5.
FT REPEAT 233 262 WD 6.
FT REPEAT 279 307 WD 7.
SQ SEQUENCE 312 AA; 34372 MW; 778321761BBCCD99 CRC64;

Query Match 4.9%; Score 81.5; DB 1; Length 312;
Best Local Similarity 23.1%; Pred.No.14;
Matches 36; Conservative 19; Mismatches 48; Indels 53; Gaps 7;

QY 53 CYSIMNGTDEGSA-WITINFLTGLSGKTPGSSVGMGLDLGGSGTQIAFLPRVEGTLOA-- 109
D 70 CVSLAHATDYALTASW-----DRSIRWDLRNGCQCKFLKHTKDVLAFAV 115
QY 110 SPGGL-----TALRMFNRTYKLSYVLGLGMSARLAILGGVGPQAKGKE--LV 160
D 116 SPDRLLIVSAGRDNRVWNVN-----VAGECMHEFLRDGHEDWVS 154

QY 161 SPCLSPSFK-----GEWEHAEVTVYRVSGQKAAASL 190
D 155 SICFSPSLEHPVIVSGSDNTIKVNVNGKCCERTL 190

RESULT 5
GBLP LEIMA STANDARD; PRT; 312 AA.
AC Q25306;
DT 15-JUL-1998 (Rel. 36, Created)
DT 18-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Guanine nucleotide-binding protein beta subunit-like protein (Antigen LACK).
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95242101; PubMed=7725103;
RA Mougneau E., Altare F., Wakil A.E., Zheng S., Coppola T., Wang Z.E., Waldmann R., Locksley R.M., Glaichenhaus N.;
RT "Expression cloning of a protective Leishmania antigen.";
RL Science 268:563-566(1995).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN BOTH STAGES OF THE PARASITE LIFE CYCLE.
CC -1- SIMILARITY: Contains 7 WD repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC ENBL; U27568; AAA97577.1; -
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINERPT.
```

```
DR PRODOM; PD000018; WD40; 4.
DR SMART; SM00320; WD40; 6.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00882; WD_REPEATS_2; 5.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat; Antigen.
FT REPEAT 9 42 WD 1.
FT REPEAT 63 93 WD 2.
FT REPEAT 105 135 WD 3.
FT REPEAT 148 180 WD 4.
FT REPEAT 192 222 WD 5.
FT REPEAT 233 262 WD 6.
FT REPEAT 279 307 WD 7.
SQ SEQUENCE 312 AA; 34413 MW; B094B566ASACID3 CRC64;

Query Match 4.9%; Score 81.5; DB 1; Length 312;
Best Local Similarity 23.1%; Pred.No.14;
Matches 36; Conservative 19; Mismatches 48; Indels 53; Gaps 7;

QY 53 CYSIMNGTDEGSA-WITINFLTGLSGKTPGSSVGMGLDLGGSGTQIAFLPRVEGTLOA-- 109
D 70 CVSLAHATDYALTASW-----DRSIRWDLRNGCQCKFLKHTKDVLAFAV 115
QY 110 SPGGL-----TALRMFNRTYKLSYVLGLGMSARLAILGGVGPQAKGKE--LV 160
D 116 SPDRLLIVSAGRDNRVWNVN-----VAGECMHEFLRDGHEDWVS 154

QY 161 SPCLSPSFK-----GEWEHAEVTVYRVSGQKAAASL 190
D 155 SICFSPSLEHPVIVSGSDNTIKVNVNGKCCERTL 190

RESULT 6
SUBD BACLI STANDARD; PRT; 274 AA.
AC P00781;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Subtilisin DY (EC 3.4.21.62).
GN APR.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE.
RC STRAIN=DY;
RX MEDLINE=84110064; PubMed=6420308;
RA Nedkov P., Oberthur W., Braunitzer G.;
RT "Primary structure of subtilisin DY.";
RL Hoppe-Seyler's Z. Physiol. Chem. 364:1537-1540(1983).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).
RC STRAIN=DY;
RX MEDLINE=99041548; PubMed=9826175;
RA Eschenburg S., Genov N., Peters K., Fittkau S., Stoeva S., Wilson K.S., Bezel C.;
RT "Crystal structure of subtilisin DY, a random mutant of subtilisin Carlsberg.";
RL Eur. J. Biochem. 257:309-318(1998).
CC -1- FUNCTION: Subtilisin is an extracellular alkaline serine protease, it catalyzes the hydrolysis of proteins and peptide amides.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity for peptide bonds, and a preference for a large uncharged residue in p1. Hydrolyzes peptide amides.
CC -1- COPACTOR: Binds 2 calcium ions per subunit.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: Secretion of subtilisin is associated with onset of sporulation, and many mutations which block sporulation at early stages affect expression levels of subtilisin. However, subtilisin is not necessary for normal sporulation.
CC -1- SIMILARITY: Belongs to peptidase family S8.
DR PDB; 1BH6; 18-NOV-98.
```

DR MEROPS, S08.037, -.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SRR; 1.
 KM Hydrolyase; Sporulation; Serine protease; Metal-binding;
 KM Calcium-binding; 3d-structure.
 FT ACT SITE 32 32 CHARGE RELAY SYSTEM.
 FT ACT SITE 63 63 CHARGE RELAY SYSTEM.
 FT ACT SITE 220 220 CHARGE RELAY SYSTEM.
 FT METAL 2 2 CALCIUM 1.
 FT METAL 41 41 CALCIUM 1.
 FT METAL 74 74 CALCIUM 1 (VIA CARBONYL OXYGEN).
 FT METAL 76 76 CALCIUM 1 (VIA CARBONYL OXYGEN).
 FT METAL 80 80 CALCIUM 1 (VIA CARBONYL OXYGEN).
 FT METAL 168 168 CALCIUM 2 (VIA CARBONYL OXYGEN).
 FT METAL 170 170 CALCIUM 2 (VIA CARBONYL OXYGEN).
 FT METAL 173 173 CALCIUM 2 (VIA CARBONYL OXYGEN).
 SQ SEQUENCE 274 AA; 27435 MW; 0154696E22F46533 CRC64;

Query Match 4.8%; Score 79.5; DB 1; Length 274;
 Best Local Similarity 22.3%; Pred. No. 18;
 Matches 70; Conservative 49; Mismatches 112; Indels 83; Gaps 18;

QY 22 GRLPLGKAKQLQKAEVFPVLDGCVSINMGDSGASAMITINFLTSLKTPG 81
 DB 7 GRLPLKAKQVQ-----AQYKGNANVKGIT-----IDGILASHT-----DLKTVG 46
 QY 82 GSSVGM-----LDLGGSSTQ-----IAPLPRVEGTLOASPPGYLTALRMFNRTYLYSASY 132
 DB 47 GASFPVSGSYNDGNGHGVAGTVALDNTTGVGVAPNLSLVAKVINS----- 98
 QY 133 LGLGMSARLALIGVEGQPADGKELVSPCL-SSEFGKEHEKAVTRYSGQVRAASLH 191
 DB 99 -GSGTYG---AIVSGIE-WATONGLDVINMSLGGS-----GSTALKQVNDKAYASGI 146
 QY 132 ELCAA---RVSEVLQNRVHREVEGVDFYAFSYYYDLAAGVLDIAEKGS---SLVVG 245
 DB 147 VVVAANSGSSGSGQNTI-----GYPAKYDVIIVGVAVDSMKRNASVSGAE 194
 QY 246 FFIANKYCRITETPOSSPSCMDLTY-----SLLOEFGFRKRYLTKRKID 296
 DB 195 LEVMAFGV-SVYSTYP-SNTYSLNGTSMASPHVAGAAALILSK--YFTLSASQVRRLS 250
 QY 297 NVETSMALGAFHY 310
 DB 251 STAIN--LGDSTY 262

RESULT 7

YQ73_VIBPA
 ID YQ73_VIBPA STANDARD; PRT; 287 AA.
 AC 087LD9;
 DT 10-OCT-2003 (rel. 42, Created)
 DT 10-OCT-2003 (rel. 42, Last sequence update)
 DE Hypothetical UPF0042 protein VP2673.
 GN VP2673.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OK NCBI_TaxID=670;
 RN (1)
 RC SEQUENCE FROM N.A.
 RP STRAIN=RIMD 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Nishimura M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 distinct from that of V. cholerae.";

RL Lancel 361:743-749 (2003).
 CC -1- SIMILARITY: Belongs to the UPF0042 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AP005082; BAC60936.1; -.
 DR HAMAP; MF 00636; -; 1.
 DR InterPro; IPR005337; UPF0042.
 DR Pfam; PF03668; ATP_bind2; 1.
 KM Hypothetical protein; ATP-binding; Complete proteome.
 FT NP_BIND 8 15 ATP (POTENTIAL).
 SQ SEQUENCE 287 AA; 32292 MW; 47B9001DA550E9A CRC64;

Query Match 4.7%; Score 78; DB 1; Length 287;
 Best Local Similarity 23.4%; Pred. No. 25;
 Matches 44; Conservative 32; Mismatches 64; Indels 48; Gaps 10;

QY 116 TALRMENRTYLYSYSLGLMSARLALIGVEGQPADGKELVSPCLSPFKGEWEHA 175
 DB 91 TLKRYSETRRIHP--LSLGGQSLSL-----DQATREKELITPLKA-----HA 132
 QY 176 EVYRYSGQKAASJHELCARVSEVLQNRVH-RTEVGVDFYAFSYYYDLAAGVLD 234
 DB 133 DLTIANSQG---SHEL-----SETVRKRVGRDRKGLVMFESEFGKYGPS-----D 178
 QY 235 AEKGSGLVWDEPIAKYCRITETPOSSPSCMDLTYVSLLOEFGFRKRYLTKRK 294
 DB 179 A-----DYFVDVRFPLNP-HWEPALRPLTGDAPIAALEQ-----HOSVLSKXQ 223
 QY 295 IDNVETSW 302
 DB 224 IESFIETW 231

RESULT 8

ENG_C_PORGI
 ID ENG_C_PORGI STANDARD; PRT; 315 AA.
 AC P59946;
 DT 15-MAR-2004 (rel. 43, Created)
 DT 15-MAR-2004 (rel. 43, Last sequence update)
 DE Probable GTPase engC (EC 3.6.1.-).
 GN ENG_C OR PG1900.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OK NCBI_TaxID=837;
 RN (1)
 RC SEQUENCE FROM N.A.
 RP STRAIN=W83;
 RX MEDLINE=22829867; PubMed=12949112;
 RA Nelson K.E., Fleischmann R.D., Deboy R.T., Paulsen I.T., Fouts D.E.,
 RA Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
 RA Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Taiton L., Gray J.,
 RA Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.U.,
 RA Dewhirst F.E., Fraser C.M.;
 RT "Complete genome sequence of the oral pathogenic bacterium
 RT Porphyromonas gingivalis strain W83.";
 RL J. Bacteriol. 185:5591-5601(2003).
 CC -1- FUNCTION: Unusual circulatory permuted GTPase that catalyzes rapid
 CC hydrolysis of GTP with a slow catalytic turnover (By similarity).
 CC -1- SUBUNIT: Monomer (Probable).
 CC -1- SIMILARITY: Contains 1 engC GTPase domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; A201718; RAQ66884.1; --
 DR TIGR; PG1900; -- 1;
 DR HAMAP; MF 01820; -- 1;
 DR PROSITE; PS00936; ENCG_GTPASE; 1;
 KW Hydrolase; GTP-binding; Complete proteome.
 FT DOMAIN 88 241 ENCG_GTPASE.
 FT NP_BIND 128 131 GTP (PROBABLE).
 FT NP_BIND 182 189 GTP (PROBABLE).
 FT NP_BIND 236 240 GTP (PROBABLE).
 FT SITE 267 290 KNUCKLE-LIKE CYSTEINE CLUSTER.
 FT SEQUENCE 315 AA; 34551 MW; FB98020A040413BE CRC64;

Query Match 4.6%; Score 77; DB 1; Length 315;
 Best Local Similarity 20.1%; Pred. No. 35;
 Matches 69; Conservative 41; Mismatches 101; Indels 132; Gaps 16;

QY 13 TPLVLKATAGRLLPGEKAQLQKVEFKASPFVLGDDCVSINMGTDGVSAMI--- 68
 Db 21 TELYMAKGNLR-----KGIRSNPTVVGDR-VEIWPASQDGGPAVKKIH 66
 QY 69 -----TINFILTSKTPGSSVGMGLDGGSTQIAFLPRVEGTLOASPPG 113
 Db 67 PRENYIIRASNLKSHILGANL-----DAAVLCTINDPTVTTFIDRLATAEAY--- 119
 QY 114 YLTALRMFNRYTKLYS-----YLGIG-----LMSAR 141
 Db 120 RVPVILVFN---KIDCYTOEDRLQLDRLSAVTAIGYPCCHVSATGEGPLDLKSLDGGK 176
 QY 142 LAILGVGEGPAKDGKELVSPCLSPFGKEWHAETVYRVSGKAAASLHELCAARVSEV 201
 Db 177 LTLAAGSGV---GKSLINALIP-----HA-----DLRTGAISQA 209
 QY 202 LQNRVHRTVEKHVDYFAFSYYVDLAAGVLIDAE--KG-GSLVVGDFPI-----AA 250
 Db 210 HHTGMMHTTFSQIMDF-----PDLSPGALIDTPGKFGTLEMGVEVSHYFPEIFAA 263
 QY 251 KYVCR-----TLETQPSQSPFSCMDLTVVSLLOE 280
 Db 264 SKGRFCNGCTHTEPGCAVLEALRRGEIAESRYISVLSILEDE 306

RESULT 9

BCHC RHOSH
 ID BCHC RHOSH STANDARD; PRT; 318 AA.
 AC Q02430; Q9RFC3;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 2-desacetyl-2-hydroxyethyl bacteriochlorophyllide A dehydrogenase
 DE (EC 1.---).
 GN BCHC
 OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
 OC Rhodobacteraceae; Rhodobacter.
 OC NCBI_TaxID=1063;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
 RX MEDLINE=93173096; PubMed=8437569;
 RA McGlynn P., Hunter C.N.;
 RT "Genetic analysis of the bchC and bchA genes of Rhodobacter
 sphaeroides.";
 RL Mol. Gen. Genet. 236:227-234 (1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
 RX MEDLINE=20115911; PubMed=10648776;
 RA Choudhary M., Kaplan S.;

RT "DNA sequence analysis of the photosynthesis region of Rhodobacter
 sphaeroides 2.4.1.";
 RL Nucleic Acids Res. 28:862-867 (2000).
 CC -I- FUNCTION: THIS PROTEIN CATALYZES THE PENULTIMATE STEP IN
 BACTERIOCHLOROPHYLL A BIOSYNTHESIS.

CC -I- PATHWAY: Light-independent bacteriochlorophyll biosynthesis.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; AJ010302; CAB38746.1; --
 DR EMBL; AF195122; AAF24296.1; --
 DR PIR; S30914; S30914.
 DR PIR; T50752; T50752.
 DR InterPro; IPR002085; Adh_zn_family.
 DR InterPro; IPR005903; BchB.
 DR Pfam; PF00107; Adh_zinc_N; 1.
 DR ProDom; PD040557; BchB_1.
 DR TIGRFAMs; TIGR01202; bchC; 1.
 KW Photosynthesis; Bacteriochlorophyll biosynthesis; Oxidoreductase.
 FT CONFLICT 278 278 A -> R (IN REF. 1).
 FT CONFLICT 301 302 GL -> AF (IN REF. 1).
 FT SEQUENCE 318 AA; 33912 MW; C9388CAB94D0E585 CRC64;

Query Match 4.6%; Score 77; DB 1; Length 318;
 Best Local Similarity 21.6%; Pred. No. 35;
 Matches 70; Conservative 40; Mismatches 94; Indels 120; Gaps 19;

QY 22 GLRLPGEKAQLQKVEFKASP-----FLVGDDCVSINMGTDGVSAMITIN 71
 Db 58 GYPLVPGVEA-----AGEVVEAAPDTGFRGDRVFVPGSNCPA---PTDAG---PIR 103
 QY 72 FLTGS-----LKTGSGSVGMGLDGGSTQIAFLPRVEGTLOASPPGYLTALRMFNRYKL 127
 Db 104 GLFAATKRLVTPAHRAV-----RIDPALEA--GALLALAAATAR-HAL 144
 QY 128 YSYSYL-----GLGLMSARLAILGVGEGPAKDGKELVSPCLSPFGKEWHAETV 178
 Db 145 AGLNHLVPLDLVGHGTLGRLLARLITAAAG--GEP-----FW-----WE----- 181
 QY 179 YRVSGQKAAASLHELCAARVSEVLQNEVHRTVEKHVDYFAFSYYVDLAAGVLIDA--- 235
 Db 182 -----TKAERRRH---AGEVEVIDPATDQRRDYRSI-----YDASGDPKLIIDSLVM 224
 QY 236 --EKGGSILVGDGDF---EIAAKYVCRTL-----ETQPSQSPFSCMDLTVVSLLOEF 281
 Db 225 RLAKGSEIVLAGFYTFEPVAFVPAFMKEARLRIAAEWQPE-----DMVATRALIESG 277
 QY 282 GFPRSKVLKTRKIDNVETSWALG 305
 Db 278 ALSLANLITTRPASEAAEAYATG 301

RESULT 10

CLN8 HUMAN
 ID CLN8 HUMAN STANDARD; PRT; 286 AA.
 AC Q9UBY8; Q96195;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE CLN8 protein.
 GN CLN8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., VARIANT EPMR GLY-24, AND VARIANT ALA-155.

RX MEDLINE=99438402; PubMed=10508524;
 RA Ranta S., Zhang Y., Roos B., Lonka L., Takkinen E., Messer A.,
 RA Sharp U., Wheeler R., Kusumi K., Mole S., Liu W., Soares M.B.,
 RA Bonaldo M.F., Hirvaentemi A., de la Chapelle A., Gilliam T.C.,
 RA Lehesjoki A.-E.;
 RT "The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice
 RT are associated with mutations in CLN8";
 RL Nat. Genet. 23:233-236(1999).
 RN [2].
 RN SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marmise K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stepien M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein W.J., Usdin T.B., Toshiyuki S., Carrino P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Ganathie P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulky S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey U., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shvachenko Y., Boulifard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smalhus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3].
 RN SUBCELLULAR LOCATION, AND MUTAGENESIS OF 283-LYS-LYS-284.
 RX MEDLINE=20320699; PubMed=10861296;
 RA Lonka L., Kyttälä A., Ranta S., Jalanko A., Lehesjoki A.-E.;
 RT "The neuronal ceroid lipofuscinosis CLN8 membrane protein is a
 RT resident of the endoplasmic reticulum";
 RL Hum. Mol. Genet. 9:1691-1697(2000).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum and ER-Golgi intermediate compartment (ERGIC).
 CC -1- PTM: Does not seem to be N-glycosylated.
 CC -1- DISEAS: Defects in CLN8 are a cause of progressive epilepsy with
 CC mental retardation (EPMR) [MIM:600143], also known as neuronal
 CC ceroid lipofuscinosis type 8 and Northern epilepsy. EPMR is an
 CC autosomal recessive disorder characterized by normal early
 CC development, onset of generalized seizures between 5 and 10 years,
 CC and subsequent progressive mental retardation. Biochemically, the
 CC disease is characterized by the intracellular accumulation of
 CC hydrophobic material, mainly ATP synthase subunit C.
 CC -1- SIMILARITY: Contains 1 TLC (TPRM/LAG3/CLN8) domain.
 CC -1- DATABASE: NAME=NCU CLN8;
 CC NOTE=Neural Ceroid Lipofuscinosis mutation db;
 CC WWW="http://www.ucl.ac.uk/nc1/CLN8.html".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC
 DR EMBL, AF123757; AAF13115.1; -
 DR EMBL, AF123758; AAF13116.1; -
 DR EMBL, AF123759; AAF13117.1; -
 DR EMBL, AF123760; AAF13118.1; -
 DR EMBL, AF123761; AAF13119.1; -
 DR EMBL, BC007725; AAH07725.1; -
 DR GeneW; HGNC:42079; CLN8.
 DR MIM; 607837; -
 DR MIM; 600143; -
 DR GO; GO:0005783; C:endoplasmic reticulum; TAS.

```

DR GO: GO:0005793; C:ER-Golgi intermediate compartment; TAS.
DR GO: GO:0016021; C:integral to membrane; TAS.
DR GO: GO:0007399; P:neurogenesis; TAS.
DR InterPro: IPR006634; TLOC.
DR SMART: SM00724; TLOC_1.
KW Transmembrane; Endoplasmic reticulum; Neuronal ceroid lipofuscinosis;
KW Disease mutation; Polymorphism; Epilepsy.
FT TRANSMEM 21 41 POTENTIAL.
FT TRANSMEM 62 84 POTENTIAL.
FT TRANSMEM 103 123 POTENTIAL.
FT TRANSMEM 131 151 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT DOMAIN 62 262 TLOC.
FT SITE 283 286 ER-RETRIEVAL SIGNAL.
FT VARIANT 24 24 R->G (in EPMR).
FT VARIANT 155 155 /FTId=VAR_013174.
FT VARIANT 155 155 V->A (in EPMR).
FT MUTAGEN 283 284 /FTId=VAR_013175.
FT CONFLICT 225 225 KK->RR: LOCALIZES TO THE GOLGI COMPLEX.
FT FT N->S (IN REF. 2).
SQ SEQUENCE 286 AA; 32842 MW; 0BAABEBCA516D1DC CRC64;

Query Match 4.6%; Score 75.5; DB 1; Length 286;
Best Local Similarity 24.5%; Pred. No. 42;
Matches 61; Conservative 28; Mismatches 87; Indels 73; Gaps 14;

QY 81 GGSSVGLMDLGGSGTQIAFLPVEGTQLQAAPPGLALRMF-----NRTYK-LYS 129
Db 7 GGTSSESLFDIDYASWG-----RSTLMVA--GVVFPLGVFWCHQSLSINATRYSLVA 58
QY 130 YSYIGLGMSRLAILGVSQAPAKSKELVPSCLSPFGEMENHAETRVVSQKLAAS 189
Db 59 REKFVMDLATRAVF--GVOSTAGIALLDGPVL-----HAD--KARGQQNWCV 104
QY 190 LH-----ELCAAVSEVLQNRHVRTEBKVHPDYAFASYDDLAAVGILDAEKGSLLV 242
Db 105 FHITTAGGFCEFEVNAVHLNSLTIRT-----FDLVIVHHLPFAIFGLF-----GCLVN 152
QY 243 --VDPEIAXKYCRPLETOPSSPFSCMDITYVSLILOEFGEFRSKYLKIKTKINDVET 300
Db 153 LQVG-----HYLAMTTLLLEMSYPFC---VSMWLLKAGMSESLFWKLNQ----- 194
QY 301 SWALGAIFH 309
Db 195 -WLIMHFPH 202

RESULT 11
V712_SYNY3 ID_V712 SYNY3 STANDARD. PRT: 322 AA.
AC Q55979;
DT 15-JUN-1998 (Rel. 36, Created)
DDT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DN Hypothetical protein slr0712.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
ON NCBI_Taxid=1148;
RN RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Sugita M., Tabata S.,
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RL region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: TO E.COLI YBHJ.
CC This SwissProt entry is copyright. It is produced through a collaboration

```


RA Fleischmann R.D., Alland D., Eiserich J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.P., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A.L., Utterback T., Weidman J., Khouli H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z70283; CAA94259.1; -;
DR EMBL; AE007072; AAK46561.1; -;
DR PIR; D70787; D70787.
DR TIGR; MT2276; -;
DR Tuberculin; Rv2219; -;
KW Hypothetical protein; Transmembrane; Signal; Complete proteome.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 250 HYPOTHETICAL PROTEIN Rv2219/MT2276.
FT TRANSMEM 51 71 POTENTIAL.
FT TRANSMEM 73 93 POTENTIAL.
FT SEQUENCE 250 AA; 26863 MW; ADC433144243095C CRC64;
CC
CC Query Match 4.5%; Score 75; DB 1; Length 250;
CC Best Local Similarity 22.1%; Pred. No. 39;
CC Matches 46; Conservative 22; Mismatches 86; Indels 54; Gaps 7;
CC
QY 72 FLTGLSKTPGSSVGMLDLGGGSGTQIAFLPRVEGTLQASPPGYLTALRPNRTYKLYSYS 131
DB 49 YMGAFLLIVGASVGVWAGGFTMTIPL--GVLL---GALVAFVIFGR----- 94
QY 132 YLGLGLMSRLAILGVGEQPAKGKELVSPCLSPSFKEWE-----HAEVTVRVS 182
DB 95 -----RAQRTVYKAEQGTGAAAWL-----DNLGKRWRTPGVAATGNLDVAHRVI 141
QY 183 GOKAASALHELCAARYSEVLQNRVHRTVEKHVDYAFVSYYDLAAGVGLIDAEKGSLV 242
DB 142 GRPGVIFVGEGAARVKPILAQEKRTAR-----LVGDVPIYDIIVGN--- 184
QY 243 VGDFEIAAKYVCRTLETQPSQSPFFSCMD 270
DB 185 -GDGEVPLAKLERHLRLPANIIVTKQMD 211
CC
RESULT 13
FPNK_AQUAE STANDARD; PRT; 274 AA.
ID ID FPNK_AQUAE
AC AC O67055;
DT 30-MAY-2000 (Rel. 39, Created)
DT 10-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probably inorganic polyphosphate/ATP-NAD kinase (EC 2.7.1.23)
DE (Poly(P)/ATP NAD Kinase).
GN FPNK OR AQ_909.
OS Aquifex aeolicus.
OC Bacteria; Aquificae;
CC NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358 (1998).
RL

```

DR EMBL; X68366; CAA48429.1; -.
DR PIR; S30305; S26440.
DR InterPro; IPR001434; DUF11.
DR Pfam; PF01345; DUF11.1.
DR TrEMBL; TIGR01451; B_ant_repeat; 1.
DR Hypothetical protein; PfamId:
KW SEQUENCE 227 AA; 24681 MW; 2F3004C25195849E CRC64;
SQ
Query Match 4.4%; Score 73.5; DB 1; Length 227;
Best Local Similarity 23.1%; Pred. No. 46;
Matches 45; Conservative 29; Mismatches 64; Indels 57; Gaps 11;
QY
Db VSAM-IITNLTGSLKTPGG-----SSVGMLDLGGSGQIARLPVEGRLQASPPQYL 115
64 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db VKANMLVQVATARVQLPEGIVVODYMGQYIDLEGTGEIGIPAYE-----ERSL 106
54 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 116 TALRMFNRYKLYSYSYGLGLMSARLAILGGVEGQPAKDGLKIVSPCLSPSEKGEWEHA 175
107 TFLCLMRT-----GSVTNANVTADGDDNANNAAEL-----TFK-VGGIS 147
Db 176 EMYRVSQGRRAASIMHELCARSEVQLNR-VHRTSEVKHVDFA-----FSY--YY 224
QY 148 DLEVNVTGNKEIARIGD--TVRIIVKLNRGPHDANNIKIGNFIUSGLVYONFSYDAGYF 205
Db 225 D-----LAAG 229
QY 206 DDIRRWIFETLLAAG 220
RESULT 15
PCNA PYRAB STANDARD; PRT; 249 AA.
AC Q9UYX8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA polymerase sliding clamp (Proliferating cell nuclear antigen
DE homolog) (PCNA).
DE PCN OR PYRAB13790 OR PAB1465.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxId=29292;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS5 / Oreay.
RX MEDLINE=22511545; PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Hellig R., Lecompte O.,
RA Poch O., Pleur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi."
RL Mol. Microbiol. 47:1495-1512(2003).
CC -!- FUNCTION: Sliding clamp subunit. Responsible for tethering the
CC catalytic subunit of DNA polymerase to DNA during high-speed
CC replication (By similarity).
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SIMILARITY: Belongs to the PCNA family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/isb-sib.ch).
CC -----
DR EMBL; AJ248287; CAB50284.1; -.
DR PIR; G75048; G75048.
DR HSSP; O73947; IGE8.
DR HAVAP; WF_003177; -.
DR InterPro; IPR000730; Pr_cel_nuc_antig.
DR Pfam; PF00705; PCNA; 1.

```

DR Pfam: PF02747; PCNA_C; 1.
DR PRINTS; PRO0339; PCNACYCLIN.
DR PRODOM; PD002673; Pr_cel_nuc_antig; 1.
DR TIGRFAMs; TIGR00590; pcna; 1.
DR PROSITE; PS01251; PCNA_1; 1.
DR PROSITE; PS00293; PCNA_2; FALSE NEG.
KW DNA-binding; DNA replication; Complete proteome.
SQ SEQUENCE 249 AA; 28035 MW; B7E21A135B3F3BC8 CRC64;

Query Match 4.4%; Score 73.5; DB 1; Length 249;
Best Local Similarity 22.0%; Pred. No. 52;
Matches 49; Conservative 31; Mismatches 100; Indels 43; Gaps 10;

QY 33 KLLQKVEYFKASPLVGGDDCVSIKNGTDEGVSAITINF---LTGSLKTPGGSVG--- 86
DB 15 QLIETASRLIDEAAFAFVTEEGIS-MRANDPSRVVLIDNLPASIFSKYEVDGEETIGVNM 73
QY 87 -----MLDLGGSTQIAFLPRVEGTLOASPPCYLTALRMENRTYK--LYSYSYLGGL-- 137
DB 74 DHLKVKLKEGKAKETILIRKGEENFLEISLQGTAT-----RTFKPLPIDVEEIEVDLPE 127
QY 138 --MSARLAILGGVEGPAKDGKELYSPCLSPFKGMEHAEVTVYVSGKAAASLHELCA 195
DB 128 LPFTAKVILGDVITKEAVXD-----ASLVSDSMKFIAKENEFTRAEG-----ETQE 174
QY 196 ARVSEVLQNR-----VHRTVEVKHVDYFAFSYYVDLAAGVGLID 234
DB 175 VEVKLTLEDEGLLDIEVQETKSA--YGISYLSDMVKGLGKAD 215

Search completed: June 8, 2004, 10:09:08
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: June 8, 2004, 10:06:02 ; Search time 45 Seconds

(without alignments)
2257,708 Million cell updates/sec

Title: US-09-905-589a-2_COPY_135_456

Perfect score: 1659
Sequence: 1 AKQDIPDFWKATPLVLKAT.....ALGAIHYIDSLNKQSPAS 322

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 651375

Minimum DB seq length: 0

Maximum DB seq length: 322

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	643.5	38.8	278	11 O8CH23	O8CH23 mus musculu
2	377	22.7	224	11 O8CDB6	O8CDB6 mus musculu
3	374.5	22.6	271	5 O8IPZ6	O8IPZ6 drosophila
4	115.5	7.0	300	11 O8K012	O8K012 mus musculu
5	113	6.8	322	16 O8G316	O8G316 bifidobacte
6	94	5.3	280	16 O69808	O69808 streptomyc
7	88	5.7	246	16 O821C9	O821C9 chlamydoma
8	86	5.2	280	16 O821R0	O821R0 streptomyc
9	85.5	5.2	244	16 O9X810	O9X810 streptomyc
10	85	5.1	322	16 O8ESU0	O8ESU0 oceanobacti
11	84	5.1	293	16 O9WZ05	O9WZ05 thermocoga
12	84	5.1	321	16 O7W077	O7W077 bordetella
13	84	5.1	321	16 O7W1V1	O7W1V1 bordetella
14	83.5	5.0	224	16 O82CJ2	O82CJ2 streptomyc
15	83.5	5.0	246	11 O8BVP6	O8BVP6 mus musculu
16	83.5	5.0	280	16 O8PAR8	O8PAR8 xanthomona

17	83	5.0	304	2 O7MWV3	O7MWV3 alcaligenes
18	82	4.9	212	2 O52738	O52738 rhodococcus
19	82	4.9	258	10 O84K39	O84K39 oryza sativ
20	82	4.9	267	10 O7XV22	O7XV22 oryza sativ
21	82	4.9	297	16 O92U54	O92U54 rhizobium m
22	82	4.9	322	5 O9XK39	O9XK39 caenorhabdi
23	81.5	4.9	267	16 O8U677	O8U677 agrobacteri
24	81.5	4.9	276	5 O43942	O43942 leishmania
25	81.5	4.9	312	5 O9BJU5	O9BJU5 leishmania
26	81.5	4.9	321	16 O7VS81	O7VS81 bordetella
27	81	4.9	183	13 O7ZUJ5	O7ZUJ5 brachydanio
28	81	4.9	223	15 O97JCT	O97JCT clostridium
29	81	4.9	279	17 O8TLR4	O8TLR4 methanobarc
30	80.5	4.9	237	3 O8XIC6	O8XIC6 talaromyces
31	80.5	4.9	312	5 O9GUB0	O9GUB0 leishmania
32	80	4.8	277	16 O81N09	O81N09 bacillus an
33	80	4.8	280	16 O821O7	O821O7 streptomyc
34	79.5	4.8	264	16 O8F7N9	O8F7N9 leptospira
35	79	4.8	322	5 O23524	O23524 caenorhabdi
36	78.5	4.7	245	17 O9YAR7	O9YAR7 aeropyrum P
37	78.5	4.7	312	5 O43944	O43944 leishmania
38	78.5	4.7	313	2 O84HP4	O84HP4 amycolatops
39	78	4.7	286	16 O8RBL2	O8RBL2 thermobacter
40	78	4.7	308	16 O8CX80	O8CX80 oceanobacti
41	77.5	4.7	176	2 O49142	O49142 methylobact
42	77.5	4.7	278	2 O9EXX2	O9EXX2 enterococcu
43	77.5	4.7	284	13 O90348	O90348 coturnix co
44	77.5	4.7	314	16 O24922	O24922 helicobacte
45	77.5	4.7	319	17 O8TY61	O8TY61 methanopyru

ALIGNMENTS

RESULT 1	ID	Q8CH23	PRELIMINARY;	PRT;	278 AA.
AC	O8CH23				
DT	01-MAR-2003	(TEMBLrel.. 23, Created)			
DT	01-JUN-2003	(TEMBLrel.. 23, Last sequence update)			
DE	Similar to ectonucleoside triphosphate diphosphohydrolase 6.				
GN	ENTPD6.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Mammary gland;				
RA	Strausberg R.;				
RU	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; BC038126; AAH38126.1; -.				
DR	MGI; MGI:1202295; Enpd6.				
DR	GO; GO:0016787; F:hydrolase activity; IEA.				
DR	InterPro; IPR000407; GDAI_CD39_NTPase.				
DR	Pfam; PF01150; GDAI_CD39; 1.				
KW	Hydrolase.				
SQ	SEQUENCE 278 AA; 30130 MW; 75A92DD1AC76297F CRC64;				
Query Match	38.8%;	Score 643.5;	DB 11;	Length 278;	
Best Local Similarity	88.1%;	Pred. No. 7.1e-50;			
Matches 126;	Conservative	7;	Mismatches 5;	Indels 5;	Gaps 1;
Oy	1	AKQDIPDFWKATPLVLKATAGRLLPGEKAGXLLQKYGVFKASPPFLVGDCCVAINMGT 60			
Db	135	AKQHIYDFWKATPLVLKATAGRLLPGEKAGXLLQKYGVFKASPPFLVGDCCVAINMGT 194			
Oy	61	DEGVSAMITINFLVGLKTPGGSSVGMIDLGGGSGTQIAFLPVEGTLQASPPGYVLTALRM 120			
Db	195	DEGVSAMITINFLVGLKTPGGSSVGMIDLGGGSGTQITFLPVEGTLQASPPGYVLTALRM 254			
Oy	121	FRTYKLVGSYVLGLGMSARLA 143			

```

Db      255 FNRTKLYSYRW-----VCSRLA 272

RESULT 2
ID Q8CDB6 PRELIMINARY; PRT; 224 AA.
AC Q8CDB6;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 6.
EN ENTPD6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK030789; BAC27138.1; -.
DR MGD; MGI:1202295; Entp6d.
DR InterPro; IPR000407; GDAL CD39_NTPase.
DR Pfam; PF01150; GDAL CD39_1.
SQ SEQUENCE 224 AA; 24538 MW; 305DCC373B5B4A8E CRC64;

Query Match 22.7%; Score 377; DB 11; Length 224;
Best Local Similarity 96.0%; Pred. No. 7.3e-26;
Matches 72; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKQIIPDFWKATPLVLKATAGLRLLPGEKAQKLQKVEFKASPFVLGDDCVSINMGT 60
135 AKQHIYDFWKATPLVLKATAGLRLLPGEKAQKLQKVEFKASPFVLGDDCVSINMGT 194
Db

QY 61 DGVGSAMITINFLTG 75
195 DGVGSAMITINFLTG 209
Db

RESULT 3
Q8IP26 PRELIMINARY; PRT; 271 AA.
AC Q8IP26;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE C3059-FC.
GN NTPASE OR C3059.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [3]
RP SEQUENCE FROM N.A.
RC MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Gabor G.L.,
RA Abiril J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

```

```

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodita C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murzyn D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinart K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.N., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC Celnikier S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Jaculi M., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Heit G., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D., Venter E.,
RA Williams S.N., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochmk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnikier S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnikier S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003591; AAN10397.1; -.
DR FlyBase; FBgn0024947; NTPase.
DR InterPro; IPR000407; GDAL CD39_NTPase.
DR Pfam; PF01150; GDAL CD39_1.
SQ SEQUENCE 271 AA; 30025 MW; BF01FF322360A363 CRC64;

Query Match 22.6%; Score 374.5; DB 5; Length 271;
Best Local Similarity 36.8%; Pred. No. 1.6e-25;
Matches 103; Conservative 41; Mismatches 101; Indels 35; Gaps 12;

```

```

QY 52 DCVSINMGTDGVSAMITINFLTSL- KTPGSSVGMIDLGGSTQIAFLPRVETGLQAS 110
DB 2 DAVEIMDGTDEGIFSMFTVNFGLRLSKT---NQAALDLGGSGTQVTFSP---TDPDQ 54
QY 111 PGY-----LTLARMERRTKYLSYSYGLGLMSARLALIGVGEPADGKEIVSPC 163
DB 55 VPVYKVAHEVVTSSKIN---VTHSYLGLMAAHANF--THGY-KKEDIVLESVC 107
QY 164 LSPSKGE-WEHAETVTVSGQKAASLHE-----LCAARY-SEVLQNRVHRTVEYKH 214
DB 108 VNPILANRTWYGNQYKYSKENGKSSAQPIVDPAELVYKSKVMPLVKPKPFTLKQ 167
QY 215 VDFVAFSYVYDLAAVGLIDAEKGSILVGDPELTAAYVCTLEHQPSGSCMDLTVY 274
DB 168 HAVAFSTYFESALISGLVDPLAGSETTEVAIRKKAQELCPNDE---QPFMCFDLFTI 224
QY 275 SLILOE-FGPPRSKVLKLRKIDNVETSMALGATFHYIDS 313
DB 225 STLREGFLNDGKKIKLYKKIDGHEISMALCAVAVLTS 264

```

RESULT 4

```

Q8K0L2 PRELIMINARY; PRT; 300 AA.
AC Q8K0L2;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DB RIKEN CDNA 2010320H07 gene.
GN 2010320H07R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC031143; AAH31143.1;
DR MGD; MGI:1919340; 2010320H07R1K.
DR InterPro; IPR000407; GDAI_CD39_NTase.
DR Pfam; PF01150; GDAI_CD39; 1.
SQ SEQUENCE 300 AA; 33645 MW; C3E372AFB266B1C CRC64;

```

Query Match 7.0%; Score 115.5; DB 11; Length 300;
 Best Local Similarity 20.7%; Pred. No. 0.056;
 Matches 59; Conservative 39; Mismatches 84; Indels 103; Gaps 12;

```

QY 85 VGMIDLGGSTQIAFLPR--VEGTQASPPGYTLALMERNRYKLSYSTLGLMSAR 141
DB 3 VGMIDLGGSTQIAFLPR--VEGTQASPPGYTLALMERNRYKLSYSTLGLMSAR 141
QY 142 LAIGVGEGOPAKXG-KELVSPCLSPSKGEWEHAETVTVSGQKAASLHE-----CAA 196
DB 52 DQILNRILAKLAQDRLSQVAPVRHPCV-----SGYQALLPLSLSDSC-- 97
QY 197 RVSEVLQNRVHRTVEYKH----- 215
DB 98 -----IHTDSINHTQNLVEGTGDPGNCVVALSLFNSCKGQKQCAFNGIYOP 148
QY 216 -----DEVAFSYVYDLAAVGLIDAEKGSILVGDPELTAAYVCTLEHQPSGSCMDLTVY 274
DB 149 PVHGFVAFSNFYTFHFLNLTSRQSLATV---NDTVWKFQCGPKVLVEVSYFGQERWL 204
QY 268 ---CMDLTVSLLQEFGRPSKV---LKLTRKIDNVETSMALG 305
DB 205 RDYCASGLYILVLLLE-GYKFSSEETWPIQFGQKQAGNWDIDWLTG 248

```

RESULT 5
 Q8G3L6 PRELIMINARY; PRT; 322 AA.
 ID Q8G3L6

```

AC Q8G3L6;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DB Widely conserved hypothetical protein.
GN B11742.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteriia; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karamantzou M., Shet B., Vilanova D., Berger B.,
RA Pessi G., Zwielen M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract."
RI Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR EMBL; AB014808; JAM25526.1;
DR InterPro; IPR003695; Ppx-GppA.
DR Pfam; PF02541; Ppx-GppA; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 322 AA; 35198 MW; 10BB35585F54BC87 CRC64;

```

Query Match 6.8%; Score 113; DB 16; Length 322;
 Best Local Similarity 24.7%; Pred. No. 0.11;
 Matches 47; Conservative 29; Mismatches 74; Indels 40; Gaps 7;

```

QY 10 WKATPVLKATAGLRLLPEKQKLLQKYKEVFKASPFVAGDVCVIMNGTDEGVSAWIT 69
DB 78 YEITQLLWMTSLRARP--NKNVYHKHE-----LIQG-VTVLSGTDEA-----R 122
QY 70 INFVLSLKTGGSSVG--MDDLGGSTQIAFLPRVETGLQASPPGYTLALMERNRYK 126
DB 123 LTFPL--NARRWYGMWAGRLVLVDIGGSLVAMGSDDEPTVALSAPA----- 167
QY 127 LVYSYSLGLMSARLALIGVGEPADGKEIVSPCLSPSKGEWEHAETVTVSGQKA 186
DB 168 -----GAGRTTELPBGMAATPDELEVRQVAKILBPMKVVPPOSGRPHAVGTSKT 220
QY 187 AASLHELCQA 196
DB 221 FSLRLCGA 230

```

RESULT 6

```

ID Q69808 PRELIMINARY; PRT; 280 AA.
AC Q69808;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DB Hypothetical protein SC06414.
GN SC06414 OR SC1A6.03.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Parkhill J., Bentley S.D., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);

```

```

RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinaishi H., Hopwood D.A.; and a detailed genomic and physical map for
RT "A set of ordered cosmids and a detailed genomic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN Mol. Microbiol. 21:77-96(1996).
RN
RN SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wierozorek A., Woodward J., Barrall B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR ENBL; AL93127; CAA18901.1; -.
DR PIR; T28684; T28684.
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.
DR GO; GO:0006807; P:nitrogen metabolism; IEA.
DR InterPro; IPR003010; Ntise/CNhydrtse.
DR Pfam; PF00785; CN hydrolase; 1.
DR PROSITE; PS50263; CN_HYDROLASE; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 280 AA; 31804 MW; CEE8483E6E40B167 CRC64;

Query Match 5.7%; Score 94; DB 16; Length 280;
Best Local Similarity 31.1%; Pred. No. 4.5;
Matches 37; Conservative 13; Mismatches 53; Indels 16; Gaps 5;

QY 147 GVGEQAKD---GKELVSPCLSPFSGEW-----EHAETVTVSGQKAAAS-----LHEL 193
DB 136 GNSGWPFVDVAVKGVYICVDHFFECWFALEGAEIVFNPSPATSGSLGYLVQLEGP 195

QY 194 CAARVSEVLQNRVHR--TEEVGHVDYAFSYVDLAAG-VGLIDAEKGLSVVGPEFA 249
DB 196 AAANVANEYFGAINRVGVEELGNDVFGTGYFVDPPEARFVGEVADSKETELVVRDLMA 254

RESULT 7
Q821C9 PRELIMINARY; PRT; 246 AA.
AC Q821C9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Virulence protein pGP8-D.
GN CCAA00001.
OS Chlamydomophila caviae.
OG Plasmid pCpGP1.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83557;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=22569155; PubMed=12682364;
RA Read T.D., Myers G.S.A., Brurham R.C., Nelson W.C., Paulsen I.T.,
RA Heidelberg J., Holtzaple E., Khouri H., Federova N.B., Carty H.A.,
RA Unayán L.A., Haft D.H., Peterson J., Beanan M.J., White O.,
RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavol P.M.,
RA Fraser C.M.;
RT "Genome sequence of Chlamydomophila caviae (Chlamydia psittaci GPIC):
RT examining the role of niche-specific genes in the evolution of the
RT Chlamydiaceae.";
RL Nucleic Acids Res. 31:2134-2147(2003).
DR ENBL; AE015926; AAP05752.1; -.
DR TIGR; CCAA00001; -.

```

```

DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR005350; UPF0137.
DR Pfam; PF03677; UPF0137; 1.
DR Plasmid; Complete proteome.
SQ SEQUENCE 246 AA; 28045 MW; 1B4C3F239179BF43 CRC64;

Query Match 5.3%; Score 88; DB 16; Length 246;
Best Local Similarity 21.1%; Pred. No. 13;
Matches 43; Conservative 77; Indels 44; Gaps 10;

QY 21 AGURLDGEKAQKLQKVEVFKASPLVGD-----DCVSINMGTDGVSANITINFLT 74
DB 65 ASIKLLTGGQ-----IKSIQKHVLLIGEKTYKREILRAVNSPDTTFSAINLVFHT 116
QY 75 GSKLTPGGSGVGMLDLGGSGTQIAFLP--RVEGTLOASPPGYLTALRMENRTYKLYSYS 132
DB 117 ---KSSAYNALGYEL-----FISLPDKTKSLFQSIP--YKTAYLLASRK----- 157
QY 133 LGGLMGAIALIGVEGQPAKQKELVSPCLSPFSGEWEHAETVTVSGQKAAASLHE 192
DB 158 ---GSVXDKVKVLGKIEGMENTAIDILNRL--PSLRASQTERSINFPEDKNKSEKLME 213
QY 193 L---CAA-RVSEVLQNRVHRTE 211
DB 214 ILKTVCSGLSELYNKLLQQLFE 237

RESULT 8
Q821LR0 PRELIMINARY; PRT; 280 AA.
AC Q821LR0;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative hydrolase.
GN SAV1950.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:528-531(2003).
RL ENBL; AP005028; BAC69661.1; -.
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.
DR GO; GO:0006807; P:nitrogen metabolism; IEA.
DR InterPro; IPR003010; Ntise/CNhydrtse.
DR Pfam; PF00795; CN hydrolase; 1.
DR PROSITE; PS50263; CN_HYDROLASE; 1.
DR Complete proteome.
SQ SEQUENCE 280 AA; 31553 MW; B25558BD9A269A43 CRC64;

Query Match 5.2%; Score 86; DB 16; Length 280;
Best Local Similarity 29.4%; Pred. No. 24;
Matches 35; Conservative 14; Mismatches 54; Indels 16; Gaps 5;

```

QY 147 GVEGPAKD---GKEIVSPCLSPSKGEME-----HAETTVRVSGCKAAS-----LHEL 193
Db 136 GNSGMPFEDTAVGIVGYYITCYDRHPFEGMRALGLAGBIVFNPATRGISGYIMQLECP 195
QY 194 CAATVEVLQNRVHR--TEEVKHYDFFAASYDYDLAAG-VGLIDAEKGSLLVGDPEIA 249
Db 196 AAANAEYFVGAINRVGVEELGDNDFYGTSTYFVDPEAFQVEVASDKETELVVRDLDLA 254

RESULT 9

Q9X8L0 PRELIMINARY; PRT; 244 AA.
ID Q9X8L0
AC Q9X8L0;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Possible membrane protein.
GN SC03431 OR SCE9.38.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteriales; Actinomycetales;
OC Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinsahl H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96 (1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)." ;
RL Nature 417:141-147 (2002).
RN EMBL, AL938116; CAB42784.1; --
DR PIR, T36357; T36357.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; F:transport; IEA.
DR InterPro; IPR000566; Lipocin cytochrome.
DR PROSITE; PS00213; LIPOCALIN_1.
KW Complete proteome.
SQ SEQUENCE 244 AA; 23480 MW; CADDBELI40F940EA CRC64;
Query Match 5.2%; Score 85.5; DB 16; Length 244;
Best Local Similarity 23.3%; Pred. No. 22;
Matches 53; Conservative 23; Mismatches 92; Indels 59; Gaps 9;
QY 67 WTIITFLNGSLKTPGGSSVGMIDLGSGSTQIAFLPRVSGTQAASPPG-----LTALRMF 121
Db 20 WVTAGNTASTENTASABESTS---GAMRTGYA-----AGTASATPAGTIVRSGGVAFRRS 70

QY 122 NRTYKLYSYSLGILGMSARLAIL-----GGVEGPAKDGEIVSPCLSPS----- 167
Db 71 RRTFRSRRVVALSLGLVGA-LALITCTGGSDGSGSGSGGSDSAAPGASATDGG 129
QY 168 -----FGEW-----EHAETTVRVSGCKA--ASLHELCAATVEVLQNRVHR 208
Db 130 DAGGSPSAGAGELGSGMLATTGQAVLMLVTDGKALFATGATVCSGTTERTSGTFRIR 189
QY 209 -----TEEVKHYDFFAASYDYDLAAG-VGLIDAEKGSLLVYG 244
Db 190 LKCADGSADPRATKXGAVGATSLTVAMGALGKETVTRSEGSLLPG 236

RESULT 10

Q9ESU0 PRELIMINARY; PRT; 322 AA.
ID Q9ESU0
AC Q9ESU0;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Thiamine monophosphate kinase (EC 2.7.4.16).
GN OB0644.
OS Oceanobacillus thelyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=22230767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus thelyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments." ;
RL Nucleic Acids Res. 30:3927-3935 (2002).
RN EMBL: AP004595; BAC12600.1; --
DR GO; GO:0009303; P:thiamin-phosphate kinase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009228; P:thiamin biosynthesis; IEA.
DR InterPro; IPR000728; AIR synth.
DR InterPro; IPR006283; ThLT.
DR Pfam; PF00586; AIRS; 1.
DR Pfam; PF02769; AIRS_C; 1.
DR TIGRFAMs; TIGR01379; thLT; 1.
KW Kinase, Transferase, Complete proteome.
SQ SEQUENCE 322 AA; 35872 MW; 6E711ADCD056B763 CRC64;

Query Match 5.1%; Score 85; DB 16; Length 322;
Best Local Similarity 21.0%; Pred. No. 36;

Matches 69; Conservative 44; Mismatches 122; Indels 94; Gaps 14;
QY 35 LQKVEVFRASPFL--VDDCVSINMGTDSEVSAWITINFTGSLKTPGGSSVGMIDLG 91
Db 7 ISIKPTVYRQPSLKGIDDDAVFRSSKDIYAVDTFEDVHFRSTWEPYHIGYRALG 66
QY 92 GSGTQIAFLPRVSGTQAASPPGYLALRM-FRITYKLYSYSLGILGMSAR-LALIGS- 147
Db 67 ANISDLA-----AMGASPSFLVSIIVPKMTQAELSQITSGDDDLASKHMDLIGD 119
QY 148 -VEG-----QPADGK-ELVSPCLSPSKG----- 170
Db 120 TVAGKQLITISIVIGYITNKARYRQHAREGDIVFVGTGLGSLAGFHILTSNENRNV 179
QY 171 -----EMEAETTVRVSGCKAASLHELCAATVEVLQNRVHRTEVKVH----- 215
Db 180 DQDFYTHRRKPEPRVAFQALBSLDRVSLNDVSGIANEASETIABASNVSTVLRNDIP 239
QY 216 ---DFAAFS---YYDLAGVGLIDAEKGSLLVGDPEIAKAYVCRTLETQPOSSPFGC 268
Db 240 VAFSFGFTYEQYQWKLSSGB---DFELLGVAKGDVAVVEAAEKT-NTQ----- 287
QY 269 MDITVYSLLOERFGR-----RSKRYLK 290


```
Db 288 --LTVIGSVVEEKHPVYIEDNAMRKVLK 314
RESULT 11
O9WZQ5
ID O9WZQ5 PRELIMINARY; PRT; 293 AA.
AC O9WZQ5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Malonyl CoA-acyl carrier protein transacylase.
GN TM0798.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
Raft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linner K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-325(1999).
DR EMBL; AB001747; AAD35880.1; -.
DR PIR; G72334; G72334.
DR HSP; P25715; IMLA.
DR TIGR; TM0798; -.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0004314; F:[acyl-carrier protein] S-malonyltransferase. .; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001227; Ac.trans.
DR InterPro; IPR004410; Fabd.
DR Pfam; PF00698; Acyl.transf. 1.
DR TIGRfams; TIGR00128; fabd; 1.
DR Complete proteome.
KW SEQUENCE 293 AA; 32966 MW; F98A4P033A88194 CRC64;
Query Match 5.1%; Score 84; DB 16; Length 293;
Best Local Similarity 24.2%; Pred. No. 39;
Matches 59; Conservative 32; Mismatches 67; Indels 86; Gaps 13;
Qy 25 LIPGEKAO-----KLQKVEFKASPFVLVGGDDCVSINMGTDGV-----SAWI 68
Db 5 VFPQGSQVGMKDFSVYESSKEIFERSKVLGFDITEIMMGDEETKLTKENAPQSIYI 64
Qy 69 TTNFLTGLSKTPGGSSVGMGLDGGGSTQIAFLPRVEGTQLQASPPGYLTALRMFRNTKLY 128
Db 65 T-----SYIAFLEL-----EKRGIIPDV---VAGHSIGETALAVAG-VYDFE 103
Qy 129 SVSYL-----GLGMSARLAILGGVGGQPAKQKELVSPCLSPFKGEW-- 172
Db 104 TGLYLVRKRGVMSKALEPGKGTMAA-----VTGLNIETIEEVN-----SIEGVIA 151
Qy 173 ---EHAEVTVRVSQ-----OKLAASHELCAARVSEVLQNRVHRT-----EEVK 213
Db 152 NYNSHQVV--ISGLKESVEKAMEILKEGARVVELVMSVSPFHTPFLEYAREKKEEVE 209
Qy 214 HVDF 217
Db 210 KVDF 213
RESULT 12
Q7WQ77
ID Q7WQ77 PRELIMINARY; PRT; 321 AA.
AC Q7WQ77;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
```

```
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative exported protein.
GN BB0238.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrett B.G., Maskell D.J.;
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640437; CAE30736.1; -.
KW Complete proteome.
KW SEQUENCE 321 AA; 33666 MW; DEF695A0B678F43E CRC64;
Query Match 5.1%; Score 84; DB 16; Length 321;
Best Local Similarity 22.8%; Pred. No. 45;
Matches 69; Conservative 37; Mismatches 136; Indels 60; Gaps 14;
Qy 17 LKATAGLRLLPGKAKOL-LQKVEFKASPFVLVGGDDCVSINMGTDGVSAWITINFLAG 75
Db 8 LLALALATPAAGAAQNWFSQPLRWIV---PTPAG-----GGTD-----VVAETVAG 51
Qy 76 SLKTPGSSVGMGLDGGGSTQIAFLPRVEGTQLQASPPGYL-----TALRMFRNTKLYSY 130
Db 52 SLEKPLGQITIVENRPGAAITIG---ATAIAQADPAGYVVGVTADSGTLAFNPSL----- 102
Qy 131 SYLGLGMSARLAILGGVGGQPAKQKELVSPCLSPFKGEWEHAEVTVRVSQKAAASL 190
Db 103 -YAKLSYDPAKFTYIGGIARFPLMLAVNNSPKYSVEDVLAARKEPAKLTRASAGAGSP 161
Qy 191 HELCAARVSEVLQNRVHRTVEVKHVDYAFS-----YYDLAGVGLIDAEK 237
Db 162 HHLAL-----ELFKQAN--VDVLHVPYKGAAPAIQDLGGQVDMFIDLAAGLPNI---K 212
Qy 238 GGSLL-VVGDFEIAAKVVCRTLETQPOSSPFCMDLTIVSL--LLQEFGEFRSKVLKLTAK 294
Db 213 AGKRLVLG---TATPERLAVLPDAPTMASQGVADFTAYAWQGLVGPAGMPEAAVKKLGGE 269
Qy 295 ID 296
Db 270 LE 271
RESULT 13
Q7W1W1
ID Q7W1W1 PRELIMINARY; PRT; 321 AA.
AC Q7W1W1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative exported protein.
GN BPP0234.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
```

RC STRAIN=12822 / ATCC BAA-587;
 RX MEDLINE=22827954; PubMed=12910271;
 RA Parkhill J., Sebastia M., Preston A., Murphy L.D., Thomson N.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Harris B., Quail M.A.,
 RA Cerdeno-Farrago A.M., Temple L., James K., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bacon N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagsels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitch E., Rutter S., Sanders M., Saunders S., Stevens K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whithead S., Barrett B.G., Maskell D.J.,
 RT "Comparative analysis of the genome sequences of *Bordetella pertussis*,
 RT *Bordetella parapertussis* and *Bordetella bronchiseptica*.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL: BX640423; CAZ39975.1; -
 KW Complete proteome.
 SQ SEQUENCE 321 AA; 33666 MW; DEF695A0B678F43E CRC64;

Query Match 5.1%; Score 84; DB 16; Length 321;
 Best Local Similarity 22.8%; Pred. No. 45;
 Matches 69; Conservative 37; Mismatches 136; Indels 60; Gaps 14;

QY 17 LKATAGRLPGEKAKL-LQKVEVFKASPELVGDDCVSINMGDGVSAWITINFLNG 75
 DB 8 LLAALALTFAAGAAQWPSQPLKMTV--PYAG-----GQTD-----VVARTVAG 51
 QY 76 SLKTPGSSVGMLDLGGSTQIAFLPRVEGTLQASPPGYL-----TALRMFNRTYKLYSY 130
 DB 52 SLEKPLGQITIVENRPGATITIG-----ATAIAQDPAGYVVGADSTLAFNLSL----- 102
 QY 131 SYLGGLMSARLALLGVGEGOPADGKELVSPCLSPFKSEWHEAVTTRVSGQKAAASL 130
 DB 103 -YAKLSYDPAKFTYIGGIARFPLMLAVVNSPYKSVEDVLQAKRKEPAKLTAASAGGSP 161
 QY 191 HELCAARVSEVLQNRVRETEVKHVDYFAS-----YYVDLAGGLDAEK 237
 DB 162 HHLML-----ELFKORAN--VDLHVPRYKGAAPALQDLGGQVDMFMFLDLAGLPLNT---K 212
 QY 238 GGSFL-VVGDPEFIARVYCRLETPQSSPSCMDLTVSL--LLOEFGFPRSKVLKLTTRK 294
 DB 213 AGKLRVIG---TAPERLAVLPDAPVMAEQGVADFVYVAGQIIVGAPGMEBAVKKLIGE 269
 QY 295 ID 296
 DB 270 LE 271

RESULT 14
 Q82CJ2 PRELIMINARY; FRT; 224 AA.

AC Q82CJ2
 DT 01-JUN-2003 (TREMblrel. 24, Created)
 DT 01-JUN-2003 (TREMblrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN SAV5353.
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21477403; PubMed=11572948;
 RX OMURA S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shimose H., Takahashi Y., Horikawa H., Nakazawa H., Oschne T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,
 RT "Genome sequence of an industrial microorganism *Streptomyces*
 RT *avermitilis*: deducing the ability of producing secondary
 RT metabolites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shimose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Omura S.,
 RT "Complete genome sequence and comparative analysis of the industrial
 RT microorganism *Streptomyces avermitilis*.";
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL: AP005042; BAC7065.1; -
 DR GO:GO:0016020; C:membrane; IEA.
 DR GO:GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
 DR InterPro: IPR003856; LPS_M2z_MPA.
 DR Pfam: PF02706; wzz; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 224 AA; 22584 MW; B7E6F5BEEF6917CC CRC64;

Query Match 5.0%; Score 83.5; DB 16; Length 224;
 Best Local Similarity 28.1%; Pred. No. 29;
 Matches 47; Conservative 15; Mismatches 80; Indels 25; Gaps 6;

QY 90 LGGSTQIAFLPRVEGTLQASPPGY-LTALRMFNRTYKLYSYGLT-----GLMSARLAI 144
 DB 33 LAAAGTLAAGLAGGAGCVL--TPPAYSATAYVIAVPTATSDPQALGLQAVGRVATGLAV 90
 QY 145 LGGVE---GQPAKDGKELVSPCLSPFKSEWHEAVTTRVSGQKRAALHLCARVSEV 201
 DB 91 LGDAQVMAGVPVKTIREVQSATSBD-----APMVAVATATSRADLATDMANVASRS 142
 QY 202 LQNRVRETEVKHVDYFASYYIDA-----AGVGLDAEKGGSL 241
 DB 143 LIRHANDIKMATHVELLQFSRAVRKAPKSASPALTGIVGASAGGL 189

RESULT 15
 Q8BVP6 PRELIMINARY; FRT; 246 AA.

AC Q8BVP6
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
 DE Hypothetical serine protease (Hypothetical protein).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; Tissue=Testis;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The PANTOM Consortium,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC Tissue=Testis;
 RA Strauberg R.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK077071; BAC36592.1; -
 DR EMBL: BC049744; AAH49744.1; -
 DR GO:GO:0004295; P:trypsin activity; IEA.
 DR GO:GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR001254; Peptidase S1.
 DR PROSITE: PS00134; TRYPSIN_HIS_1.
 KW Hypothetical protein.
 SQ SEQUENCE 246 AA; 26996 MW; 9D5FA13F10CC79D CRC64;

Query Match 5.0%; Score 83.5; DB 11; Length 246;
 Best Local Similarity 21.8%; Pred. No. 33;
 Matches 53; Conservative 36; Mismatches 89; Indels 65; Gaps 11;

QY 10 WKATPLVLTATGRLPGEKAKL-LQKVEVFKASPELVGDDCVSINMGDGVSAWIT 69

Db	6	WRSLOLLYLEA-ISLPCTEALLCYEATASAFRA-----VSLHNW-----KWLL	49
QY	70	I-----NFLTGLKT-----PGSSVGMLDLGGSTQIAFLPRVEGTLOASPPGYLTALRM	120
Db	50	LRSMVCNOREGCEETVFIETGTSKVLSPKGCSSAFSPYQI--SYLVSPFGVSI-----	104
QY	121	FNRTYKLYSYSLGLGL-----MSARLAILGGVEGOPAKDGKE-----	158
Db	105	---SYSRVCRSYLCNNLTNLEFFVRLKASQPMSTLPSAKSCPSCVGKHDOECLPSFVYTE	161
QY	159	-----LVSPCLSPSPKGEWEHAETVRVG-----OKAAASLHELCAARVSEVLQNRVHRT	209
Db	162	NCPFAASSCYSTLKFQAGNLNTTFLMGCARDSHKLLADFOHIGSIRTEVI-NVLDKS	220
QY	210	EVV	212
Db	221	EAV	223

Search completed: June 8, 2004, 10:10:06
Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 10:02:16 ; Search time 58 Seconds
(without alignments)
1568.626 Million cell updates/sec

Title: US-09-905-589a-2_COPY_135_456
Perfect score: 1659
Sequence: 1 AKQIDPFDFWRATPLVLKAT.....ALGALFHYIDSLNRQKSPAS 322

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1278565

Minimum DB seq length: 0
Maximum DB seq length: 322

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1980s:*
3: geneseqp2000s:*
4: geneseqp2000s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	596	35.9	141	AAU30879	Novel hum
2	596	35.9	144	AAU30881	Novel hum
3	189.5	11.4	243	AAQ22051	Arabidops
4	186	11.2	139	AAU30880	Novel hum
5	184	11.1	217	AAQ22052	Arabidops
6	151.5	9.1	269	AAE01419	Human gen
7	151.5	9.1	288	AAE01356	Human gen
8	151.5	9.1	288	ABG64123	Human alb
9	113	6.8	322	ABP66314	Bifidobac
10	104	6.3	46	AAU44306	Peptide #
11	104	6.3	46	ABR33253	Peptide #
12	104	6.3	46	AAU26716	Peptide #
13	104	6.3	46	ABR28079	Human pep
14	104	6.3	46	ABR18715	Protein #
15	104	6.3	46	AAU66436	Human Don
16	104	6.3	46	AAU54045	Human bra
17	104	6.3	46	ABG48102	Human liv
18	104	6.3	46	AAU02035	Peptide #
19	104	6.3	46	ABR36085	Human pep
20	86	5.2	279	ABR59553	Drosophi
21	86	5.1	281	ABR67409	Abbs7409
22	84.5	5.1	274	AAU14887	Subtilisi
23	84.5	5.1	274	AAU14866	Subtilisi
24	84.5	5.1	274	AAU14874	Subtilisi
25	83.5	5.0	274	AAU14880	Subtilisi

26	83	5.0	212	2	AAU14161	AAU14161 R. rhodoc
27	83	5.0	212	2	AAU02292	Beta chal
28	83	5.0	293	2	AAU28961	AAU28961 Serratia p
29	82.5	5.0	192	5	ABP63032	Abp63032 Human pol
30	82.5	5.0	198	5	ABP41815	Abp41815 Human ova
31	82.5	5.0	274	2	AAU14886	AAU14886 Subtilisi
32	81.5	4.9	243	4	ABG01294	Novel hum
33	81.5	4.9	274	2	AAU14868	Subtilisi
34	81.5	4.9	283	6	ABP57689	ABP57689 Sacccharop
35	81.5	4.9	299	7	ADBE4638	ADBE4638 Human pro
36	81	4.9	212	2	AAU13898	AAU13898 Nitrite h
37	81	4.9	293	2	AAU07057	Serratia p
38	80.5	4.9	274	1	AAU80849	Sequence
39	80.5	4.9	274	2	AAU14872	Subtilisi
40	80.5	4.9	274	2	AAU14878	Subtilisi
41	80.5	4.9	274	2	AAU14854	Subtilisi
42	80.5	4.9	274	2	AAU14863	Subtilisi
43	80.5	4.9	274	2	AAU14871	Subtilisi
44	80.5	4.9	274	2	AAU14851	Subtilisi
45	80.5	4.9	274	2	AAU14856	Subtilisi

ALIGNMENTS

RESULT 1	AAU30879	AAU30879 standard; protein, 141 AA.
ID	AAU30879;	
XX	AAU30879;	
XX	18-DEC-2001 (first entry)	
DT		
XX	Novel human secreted protein #1370.	
DB		
XX	Human; vaccination; gene therapy; nutritional supplement;	
KW	stem cell proliferation; haematopoiesis; nerve tissue regeneration;	
KW	immune suppression; immune stimulation; anti-inflammatory; leukaemia.	
XX		
OS	Homo sapiens.	
XX		
XX	WO200179449-A2.	
PD	25-OCT-2001.	
XX		
XX	16-APR-2001; 2001WO-US008656.	
PR	18-APR-2000; 2000US-00552929.	
PR	26-JAN-2001; 2001US-00770160.	
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Tang YT, Liu C, Drmanac RT;	
DR	WPI; 2001-611725/70.	
PT		
PT	Nucleic acids encoding a range of human polypeptides, useful in genetic	
PT	vaccination, testing and therapy.	
XX		
XX	Claim 20; Page 365; 765pp; English.	
PS		
XX	The invention relates to novel human secreted polypeptides. The	
XX	polypeptides and antibodies to the polypeptides are useful for	
CC	determining the presence of or predisposition to a disease associated	
CC	with altered levels of polypeptide. The polypeptides are also useful for	
CC	identifying agents (agonists and antagonists) that bind to them. Cells	
CC	expressing the proteins are useful for identifying a therapeutic agent	
CC	for use in treatment of a pathology related to aberrant expression or	
CC	physiological interactions of the polypeptide. Vectors comprising the	
CC	nucleic acids encoding the polypeptides and cells genetically engineered	
CC	to express them are also useful for producing the proteins. The proteins	
CC	are useful in genetic vaccination, testing and therapy, and can be used	
CC	as nutritional supplements. They may be used to increase stem cell	

CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
XX
SQ Sequence 141 AA;

Query Match 35.9%; Score 596; DB 4; Length 141;
Best Local Similarity 77.1%; Pred. No. 1.7e-52;
Matches 121; Conservative 2; Mismatches 14; Indels 20; Gaps 1;

Qy 163 CLSPFSGEWEHAEVTVYRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVYKVDYFAFSY 222
Db 5 CLSPAF-----ILLELCAARVKEGLPNRVHRTTEVYKVDYFAFSY 44

Qy 223 YYDLAAGVGLIDAEKGSLLVVGDFEATAKVCRTLETQPOSSPFSCMDLTVYSLLLQREG 282
Db 45 YYDLAAGAGPIDAEKGSLLVVGDFEATKVCRTLETQPOSSPFSCMDLTVYSLLLQREG 104

Qy 283 FPRSKVLKLRKIDNVETSWALGAIFHYIDSLNRQKS 319
Db 105 FPRSKVLKLRKIDNVETSWALGAIFHYIDSLNRQKS 141

RESULT 2
AAU30881
ID AAU30881 standard; protein; 144 AA.
XX
AC AAU30881;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #1372.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
PS Claim 20; Page 365; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell

CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
XX
SQ Sequence 144 AA;

Query Match 35.9%; Score 596; DB 4; Length 144;
Best Local Similarity 76.2%; Pred. No. 1.8e-52;
Matches 122; Conservative 3; Mismatches 1; Indels 34; Gaps 2;

Qy 163 CLSPFSGEWEHAEVTVYRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVYKVDYFAFSY 222
Db 5 CLAP-----AASLHELCAAKVSEVLHNRVHRTTEVYKVDYFAFSY 44

Qy 223 YYDLAAGVGLIDAEKGSLLVVGDFEIAAKY-----VCRTLETQPOSSPFSC 268
Db 45 YYDLAAGVGLIDAEKGSLLVVGDFEIAAKYGVVTSVYKGRVSSPVCRTLETQPOSSPFSC 104

Qy 269 MDLTVVSLLLQEFQFPRSKVLKLRKIDNVETSWALGAIF 308
Db 105 MDLTVVSLLLQEFQFPRSKVLKLRKIDNVETSWALGAIF 144

RESULT 3
AAG22051
ID AAG22051 standard; protein; 243 AA.
XX
AC AAG22051;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 24833.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125789P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 08-APR-1999; 99US-0128234P.
PR 16-APR-1999; 99US-0128714P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130691P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.

PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137520P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140698P.
PR 28-JUN-1999; 99US-0140823P.
PR 30-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.

PR 27-JUL-1999; 99US-0145818P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151388P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155569P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156599P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158929P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.

```
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161932P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 11.4%; Score 189.5; DB 3; Length 243;
Best Local Similarity 24.7%; Pred. No. 1.5e-10;
Matches 60; Conservative 43; Mismatches 91; Indels 49; Gaps 8;

Qy 103 VEGTLQSPGGLTALRMFNRTYKLYSYLGLGLMSARLILGGVGPAXDGKELVSP 162
Db 17 VEG-----ESYVREYVYKGRKFLVHSYLHYGLAARAIL-----KVSDESNP 63

Qy 163 CLSPFSGEWEHAEVTVRSVGKAAASLHELCAARVSEVLQNRVHRTTEVKHY----- 215
Db 64 CIATGYAGTYKYGKAFKAAASPSGASLDCRRVAINALKVNSLCT---HMKCTFGGV 119

Qy 216 -----DFYAFSYVYDLAAGVGLIDAEKGSVVG-DEIRAKYVCRILETQPS 264
Db 120 WNGGGGGGQKMFVASFPPDRAAEAGFVDNPQVAEVRPLDPEKAANKACNRMMEGRSK 179

Qy 265 -----PFSQMDLTY-VSLLLQBFGRPSKVLKLRKID-----NVETSWALGAIFHY 310
Db 180 FPRVEEDNLFVCLDLVYQVTLVLDVGFGLKPSQITILVKVKYGDYVAEAAWPLGSAIEA 239

Qy 311 IDS 313
Db 240 VSS 242

RESULT 4
AAU30880
ID AAU30880 standard; protein; 139 AA.
AC AAU30880;
XX
XX
XX 18-DEC-2001 (first entry)
XX Novel human secreted protein #1371.
XX
XX Human; vaccination; gene therapy; nutritional supplement;
XX stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
XX Homo sapiens.
XX
XX WO200179449-A2.
XX
XX 25-OCT-2001.
XX
XX 16-APR-2001; 2001WO-US008656.
XX
XX 18-APR-2000; 2000US-00552929.
XX
XX 26-JAN-2001; 2001US-00770160.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-611725/70.
XX
XX Nucleic acids encoding a range of human polypeptides, useful in genetic
XX vaccination, testing and therapy.
XX
XX Claim 20; Page 365; 765pp; English.
XX
XX The invention relates to novel human secreted polypeptides. The
XX polypeptides and antibodies to the polypeptides are useful for
XX determining the presence of or predisposition to a disease associated
```

```
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
XX Sequence 139 AA;

Query Match 11.2%; Score 186; DB 4; Length 139;
Best Local Similarity 75.0%; Pred. No. 1.4e-10;
Matches 39; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 1 AKQDIPDFWKATPLVKKATAGLLELPGEKAQKLLQKVKESFELVGGD 52
Db 87 AKQDIPDFWKATPLVKKATAGLLELPGEKAQKLLQKVKESFELVGGD 139

RESULT 5
AAU22052
ID AAU22052 standard; protein; 217 AA.
XX
XX AAG22052;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 24834.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 28-APR-1999; 99US-0130891P.
XX 30-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-0132407P.
XX 04-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 07-MAY-1999; 99US-0132487P.
XX 07-MAY-1999; 99US-0132863P.
XX 11-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
```

PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 14-JUN-1999; 99US-0138847P.
PR 16-JUN-1999; 99US-0139119P.
PR 17-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140911P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144633P.
PR 21-JUL-1999; 99US-0144864P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.

PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147202P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148584P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 22-AUG-1999; 99US-0149902P.
PR 22-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151803P.
PR 31-AUG-1999; 99US-0151838P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158332P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158993P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159538P.
PR 21-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160982P.
PR 25-OCT-1999; 99US-0161404P.

PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161922P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.

Query Match 11.1%; Score 184; DB 3; Length 217;
Best Local Similarity 25.1%; Pred. No. 4.5e-10;
Matches 56; Conservative 40; Mismatches 83; Indels 44; Gaps 7;

QY	123 RYKLYSYNYLGLMSRLAILGVGGQPAPAKGKELVSPCLSPFGKGWEHAETVRS	182
DB	6 RKYFLYHSYLHYGLLAARAIL-----KYSEDSNNFCIATGYAGTYKYGKAFKA	57
QY	183 GKGAASAHSELCAARVSVLQRVHRTVEVKHV-----DPAFASYKD	225
DB	58 APSGSALDECERVAINALKVNSLCI---HKCTFGGWNGGGGQKMVFASFDD	113
QY	226 LAAGVGLIDAKGSLVG-DPEIAKYVCRTLETPQSS-----PFSCMDLT-Y-V	274
DB	114 RAEEAGFDPNQPVAEVRPLDFEKAANKACNMRECKSKFRVEEDNLPYLCLDLVQY	173
QY	275 SILLQEFGRPSKVLKTRKID----NVETSWALGAIFHYIDS	313
DB	174 TLLVDGFLGKFSQITLLVKRYGDYAVEAANPLGSAIEAVSS	216

RESULT 6

AAE01419

ID AAE01419 standard; protein; 269 AA.

XX AAE01419;

XX XX

DT 18-JUL-2001 (first entry)

XX XX

DE Human gene 5 encoded secreted protein fragment, SEQ ID NO:143.

XX XX

KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma; foetal abnormality; developmental abnormality; haematopoietic disorder; immune system disease; AIDS; autoimmune disease; rheumatoid arthritis; Parkinson's disease; cognitive disorder; schizophrenia; skin disorder; psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder; inflammation; neurological disorder; Alzheimer's disease; food additive; angigenic disorder; kidney disorder; gastrointestinal disorder; allergy; pregnancy-related disorder; endocrine disorder; infection; wound healing; cell culture; chemotaxis; vulnery; binding partner identification;

KW KW

OS Homo sapiens.

XX XX

PH Key Location/Qualifiers

FT Misc-difference 87 /note= "Corresponds to any of the naturally occurring L-amino acids"

FT FT

FT Misc-difference 99 /note= "Corresponds to any of the naturally occurring L-amino acids"

FT FT

FT Misc-difference 230 /note= "Corresponds to any of the naturally occurring L-amino acids"

FT FT

FT Misc-difference 263 /note= "Corresponds to any of the naturally occurring L-amino acids"

FT FT

FT Misc-difference 264 /note= "Corresponds to any of the naturally occurring L-amino acids"

FT FT

XX WO200134629-A1.

PX XX

PD	XX		17-MAY-2001.
XX	XX		08-NOV-2000; 200OWO-USO30654.
XX	PR	PF	12-NOV-1999; 99US-0164835P.
XX	PR	PR	27-JUL-2000; 200OUS-0221142P.
XX	XX		(HUMA-) HUMAN GENOME SCI INC.
XX	XX		Ruben SM, Komatsoulis GA, Wei P, Fiscella M, Baker KP;
PI	DR		WPI; 2001-308779/32.
XX	XX		New nucleic acid encoding one of 21 human secreted proteins for
PT	PT		diagnosing, preventing, treating or ameliorating medical conditions, such
PT	PT		as autoimmune disease and cancer, and used as a food additive or
PT	XX		preservative.
XX	XX		Disclosure; Page 469-470; 490pp; English.
XX	XX		AAD05220-AAD05282 represent cDNAs corresponding to 21 human secreted
CC	CC		protein genes, and AAE01352-AAE01413 represent the proteins they encode.
CC	CC		AAE01413-AAE01433 represent human secreted protein fragments or variants.
CC	CC		The secreted proteins and their genes are useful for preventing, treating,
CC	CC		or ameliorating medical conditions, e.g., by protein or gene therapy.
CC	CC		Pathological conditions can be diagnosed by determining the amount of the
CC	CC		new protein in a sample or by determining the presence of mutations in
CC	CC		the new genes. Specific uses are described for each of the 21 genes,
CC	CC		based on the tissues in which they are most highly expressed, and include
CC	CC		developing products for the diagnosis or treatment of proliferative
CC	CC		disorders, cancer, tumours, foetal and developmental abnormalities,
CC	CC		haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC	CC		diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC	CC		neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC	CC		cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC	CC		psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC	CC		apoptogenic disorders, kidney disorders, gastrointestinal disorders,
CC	CC		pregnancy-related disorders, endocrine disorders, and infections. The
CC	CC		proteins can also be used to aid wound healing and epithelial cell
CC	CC		proliferation, to prevent skin aging due to sunburn, to maintain organs
CC	CC		before transplantation, for supporting cell culture of primary tissues,
CC	CC		to regenerate tissues, to identify their cognate ligands or binding
CC	CC		partners, and in chemotaxis, and can be used as a food additive or
CC	CC		preservative to modify storage properties. Antibodies specific for a
CC	CC		protein of the invention can be used in alleviating symptoms associated
CC	CC		with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC	CC		radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
CC	CC		present sequence represents a human secreted protein fragment referred to
CC	CC		in the disclosure of the invention
XX	XX		
SQ			Sequence 269 AA;
			Query Match 9.1%; Score 151.5; DB 4; Length 269;
			Best Local Similarity 35.1%; Pred. No. 1.3e-06;
			Matches 40; Conservative 14; Mismatches 45; Indels 15; Gaps 2
Qy	1	AKQDIPFFWKATPVILKATAGLILLGEKAOKLL-QKUKEVKASPFPLVGDDCVSIWNG	59
		:::::	:::::
Dd	155	AAARVPVKKHETPYILICTAGMTLLPERLKQLALDLVKKDLPEDFLFQSQAQEVISG	214
		:::::	:::::
Qy	60	TDEGVSAWTTTNFL-----TGLSKTPGGSSVGMLDLGGGTQTAF	99
		:::::	:::::
Dd	215	KQEGYANIGINFLVRFDHEDESDAETOLAAGRRTTVGIIDMGAXXQJAY	268
		:::::	:::::
RESULT 7			
ID	AAE01356		
AC	AAE01356 standard; protein; 288 AA.		
XX	AAE01356;		
DT	18-JUL-2001 (first entry)		
XX	XX		

DE Human gene 5 encoded secreted protein HDPH52, SEQ ID NO:78.
 XX
 KM Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
 KM foetal abnormality; developmental abnormality; haematopoietic disorder;
 KM immune system disease; AIDS; autoimmune disease; rheumatoid arthritis;
 KM Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
 KM psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
 KM inflammation; neurological disorder; Alzheimer's disease; food additive;
 KM angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
 KM pregnancy-related disorder; endocrine disorder; infection; wound healing;
 KM cell culture; chemotaxis; vulnerability; binding partner identification;
 KM gene therapy.
 KM
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /label= Signal_peptide
 FT Protein 22..288
 FT /label= Mature_human_secreted_protein
 FT Misc-difference 87
 FT /label= Unknown
 FT /note= "Encoded by GKT"
 FT Misc-difference 99
 FT /label= Unknown
 FT /note= "Encoded by GKT"
 FT Misc-difference 230
 FT /label= Unknown
 FT /note= "Encoded by SGA"
 FT Misc-difference 263
 FT /label= Unknown
 FT /note= "Encoded by YCT"
 FT Misc-difference 264
 FT /label= Unknown
 FT /note= "Encoded by CYC"
 FT Misc-difference 270
 FT /label= Unknown
 FT /note= "Encoded by GNT"
 FT Misc-difference 275
 FT /label= Unknown
 FT /note= "Encoded by NNA"
 FT WO200134629-A1.
 PN 17-MAY-2001.
 XX
 XX 08-NOV-2000; 2000WO-US030654.
 PF 12-NOV-1999; 99US-0164835P.
 PR 27-JUL-2000; 2000US-0221142P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Ruben SM, Komatsoulis GA, Wei P, Fiscella M, Baker KP;
 PI WPI; 2001-308779/32.
 DR N-PSDB; AAD05224.
 XX
 XX New nucleic acid encoding one of 21 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions, such
 PT as autoimmune disease and cancer, and used as a food additive or
 PT preservative.
 PS
 PS Claim 11; Page 436-437; 490pp; English.
 XX
 XX AAD05220-AAD05282 represent cDNAs corresponding to 21 human secreted
 CC protein genes, and ABE01352-ABE01413 represent the proteins they encode.
 CC AAB01415-AAB01433 represent human secreted protein fragments or variants.
 CC The secreted proteins and their genes are useful for preventing, treating
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.
 CC Pathological conditions can be diagnosed by determining the amount of the
 CC new protein in a sample or by determining the presence of mutations in
 CC the new genes. Specific uses are described for each of the 21 genes,

CC based on the tissues in which they are most highly expressed, and include
 CC developing products for the diagnosis or treatment of proliferative
 CC disorders; cancer, tumours, foetal and developmental abnormalities,
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,
 CC pregnancy-related disorders, endocrine disorders, and infections. The
 CC proteins can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, for supporting cell culture of primary tissues,
 CC to regenerate tissues, to identify their cognate ligands or binding
 CC partners, and in chemotaxis, and can be used as a food additive or
 CC preservative to modify storage properties. Antibodies specific for a
 CC protein of the invention can be used in alleviating symptoms associated
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
 CC present sequence represents a human secreted protein of the invention
 XX
 XX Sequence 288 AA;
 SQ
 Query Match 9.1%; Score 151.5; DB 4; Length 288;
 Best Local Similarity 35.1%; Pred. No. 1,4e-06;
 Matches 40; Conservative 14; Mismatches 45; Indels 15; Gaps 2;
 Qy 1 AKODIPDFPKATPVLYKATGAPRLPGEKATLL-QKVEVPRKASPFVGDGCVSIMG 59
 Db 155 AAHAVPKGKRPVLYLCTNKGRLPERKQALADYVDLPEDFLFSGQAVISG 214
 Qy 60 TDEGSAMITINFL-----TGLKTPGSSVGMIDLGSGTQIAF 99
 Db 215 KQSGVAMIGINFLVLRFDHEDSDAHTQELAAGRRTVGIIDMGAXXQZAY 268
 RESULT 8
 ID ABE64123 standard; protein; 288 AA.
 XX
 AC ABE64123;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 XX Human albumin fusion protein #798.
 DE
 XX
 XX Albumin fusion protein; therapeutic protein X; human albumin; HA;
 KM human serum albumin; HSA; cancer; reproductive disorder;
 KM digestive disorder; immune disorder; endocrine disorder;
 KM haematopoietic disorder; neural disorder; connective disorder;
 KM cytoskeletal; antiinfectivity; antiinflammatory; antitumor;
 KM immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
 KM neuroprotective; antiParkinsonian; antimicrobial; neuroleptic;
 KM osteopathic; antiarthritic.
 XX
 XX Homo sapiens.
 OS Synthetic.
 OS
 XX W0200177137-A1.
 EN
 XX 18-OCT-2001.
 PD
 XX 12-APR-2001; 2001WO-US011988.
 PF
 XX 12-APR-2000; 2000US-0229358P.
 PR 23-APR-2000; 2000US-019584P.
 PR 21-DEC-2000; 2000US-0256931P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Haseltine WA;
 PI WPI; 2002-010886/01.
 DR

XX New fusion protein for treating disease e.g. diabetes comprises an
PT albumin fused to a therapeutic protein.
XX
XX Claim 1; Page 1026-1027; 2102pp; English.
XX
CC The present invention relates to albumin fusion proteins comprising a
CC therapeutic protein X and human albumin (HA), also known as human serum
CC albumin, HSA). The proteins are useful for treating a disease or disorder
CC that may be modulated by therapeutic protein X. The albumin extends the
CC shelf-life of protein X, and may increase its biological in vitro/in vivo
CC activity. The protein is useful for treating and diagnosing disorders
CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's
CC disease, ulcerative colitis), immune disorders (e.g. acquired
CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),
CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,
CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,
CC schizophrenia) and connective disorders (e.g. osteoporosis, arthritis).
CC ABG63326-ABG65518 represent albumin fusion proteins of the invention
XX
SQ Sequence 288 AA;
Query Match 9.1%; Score 151.5; DB 5; Length 288;
Best Local Similarity 35.1%; Pred. No. 1.4e-06;
Matches 40; Conservative 14; Mismatches 45; Indels 15; Gaps 2;
Qy 1 AKQIPDFWKATPLVLKATAGRLLPGEKAQKLL-QKVEVKFASFLVGDGCVSIMNG 59
Db 155 AAHPVVKCHETPLYLTCTAGMRLLPKQKQAILADLVKDLPLEDFLFSQSQAQVIG 214
Qy 60 TDEGVSAWITNFI-----TGSLKTPGGSSVGMLDLGGSGTQIAF 99
Db 215 KQGVYAWIGVNFVLRPDHEDSDAEATQELAGRRRTVGILDMGAXXQIAY 268
RESULT 9
ABP66314
ID ABP66314 standard; protein; 322 AA.
XX
AC ABP66314;
XX
DT 19-NOV-2002 (first entry)
XX
DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:1058.
XX
KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KW antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KW rotavirus; food composition; pharmaceutical composition.
XX
OS Bifidobacterium longum.
XX
EN EP1227152-A1.
XX
PD 31-JUL-2002.
XX
PF 30-JAN-2001; 2001EP-00102050.
XX
PR 30-JAN-2001; 2001EP-00102050.
XX
PA (NEST) SOC PROD NESTLE SA.
XX
PA WPI; 2002-668397/72.
XX
XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as
PT a probe or primer for detecting and/or identifying Bifidobacterium longum
PT in a biological sample.
XX
XX Claim 3; SEQ ID NO 1058; 80pp; English.
XX
XX The present invention describes a polynucleotide (I) comprising a
CC sequence of a Bifidobacterium genome selected from the nucleotide
CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at

CC least 90% identity or which hybridises with the sequences given in
CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a
CC fusion protein, comprising a sequence selected from 1097 sequences given
CC in ABP65258 to ABP66354 ligated in frame to a polynucleotide encoding a
CC heterologous polypeptide. (I) has antidiarrheic and antibacterial
CC activities, and can be used as an inhibitor of Salmonella. (II) which is
CC a probe) is useful for the detection and/or identification of
CC Bifidobacterium longum in a biological sample. A carrier containing the
CC lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM 1-2618) can be
CC used for preventing and/or treating diarrhoea brought about by pathogenic
CC bacteria and/or rotavirus. The carrier is a food composition selected
CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented
CC products, ice-cream, fermented cereal based products, milk based
CC powders, infant formula, pet food or a pharmaceutical composition
CC selected from tablets, liquid bacterial suspensions, dried oral
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC (I) is useful in DNA arrays or chips to carry out analysis of the
CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC Bifidobacterium related nucleotide sequences given in the sequence
CC listing from the present invention but not mentioned further within the
CC specification. N.B. The sequence data for this patent is not represented
CC in the printed specification but is based on sequence information
CC supplied by the European Patent Office
XX
SQ Sequence 322 AA;
Query Match 6.8%; Score 113; DB 5; Length 322;
Best Local Similarity 24.7%; Pred. No. 0.015;
Matches 47; Conservative 29; Mismatches 74; Indels 40; Gaps 7;
Qy 10 WKATPLVLKATAGRLLPGEKAQKLLQKVEVKFASFLVGDGCVSIMNGDECVSAWIT 69
Db 78 YEITQLLWATSALREAP--NGNKVIHKEE-----LIQGG-VTVLSGTDEA-----R 122
Qy 70 INFITGSLKTPGGSSVG---MLDLGGSGTQIAFLPRVEGTLOASPQGYLTALRMFNRTYK 126
Db 123 LTFL--AARRWYGDAGRLLVLDIGGSGLEVAMGSDDEPTVALSAPA----- 167
Qy 127 LYSYSLGLGLMSARLAILGGVEGQPAKDGKELVSPCLSPSEKGEWHAETVTVRVSGQKA 186
Db 168 -----GAGRTTEFLPSGNATPDELDVKNVRKILEPMVKVPQSKPHNHAVGTSKT 220
Qy 187 AASLHELCAA 196
Db 221 FRSLARLCOGA 230
RESULT 10
AAM14306
ID AAM14306 standard; protein; 46 AA.
XX
AC AAM14306;
XX
DT 12-OCT-2001 (first entry)
XX
DE Peptide #740 encoded by probe for measuring cervical gene expression.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
XX 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000670.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.

27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI
XX WPI; 2001-488901/53.
XX DR
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX
XX
XX Claim 27; SEQ ID NO 19132; 487pp; English.
XX PS
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENSE: see AAI10068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 46 AA;
SQ
Query Match 6.3%; Score 104; DB 4; Length 46;
Best Local Similarity 47.6%; Pred. No. 0.0068;
Matches 20; Conservative 7; Mismatches 15; Indels 0; Gaps 0;
QY 192 ELCAARVSEVLQNRVHRTSEVKHVDYAFSYYYDLAAGVLI 233
DB 5 EPCYAEVLRVVRGKLMQPEEVQSGSFYAFSYTDRVDTDMI 46
RESULT 11
ID ABB33253
XX ABB33253 standard; peptide; 46 AA.
XX AC
XX ABB33253;
XX
XX 04-FEB-2002 (first entry)
XX DE Peptide #759 encoded by human foetal liver single exon probe.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX KM
XX Homo sapiens.
XX OS
XX WO200157277-A2.
XX PN
XX 09-AUG-2001.
XX PD
XX 30-JAN-2001; 2001WO-US000669.
XX PF
XX 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI
XX WPI; 2001-483447/52.
XX DR
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human fetal liver.
XX

Claim 27; SEQ ID NO 25888; 639pp + Sequence Listing; English.
XX PS
XX
XX The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 46 AA;
SQ
Query Match 6.3%; Score 104; DB 4; Length 46;
Best Local Similarity 47.6%; Pred. No. 0.0068;
Matches 20; Conservative 7; Mismatches 15; Indels 0; Gaps 0;
QY 192 ELCAARVSEVLQNRVHRTSEVKHVDYAFSYYYDLAAGVLI 233
DB 5 EPCYAEVLRVVRGKLMQPEEVQSGSFYAFSYTDRVDTDMI 46
RESULT 12
ID AAM26716
XX AAM26716 standard; protein; 46 AA.
XX AC
XX AAM26716;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #753 encoded by probe for measuring placental gene expression.
XX KM Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX OS
XX Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI
XX WPI; 2001-488897/53.
XX DR
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human placenta.
XX
XX Claim 27; SEQ ID NO 26985; 654pp; English.
XX PS
XX
XX The present invention relates to single exon nucleic acid probes (SENPs:
CC see AAI3315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders
XX
XX Sequence 46 AA;
SQ
Query Match 6.3%; Score 104; DB 4; Length 46;
Best Local Similarity 47.6%; Pred. No. 0.0068;


```
XX
AC AAM66436;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 26742.
XX
KM Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000668.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX
PS Example 4; SEQ ID NO 26742; 658bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention
XX
SQ Sequence 46 AA;
XX
Query Match 5.3%; Score 104; DB 4; Length 46;
Best Local Similarity 47.6%; Pred.No. 0.0068;
Matches 20; Conservative 7; Mismatches 15; Indels 0; Gaps 0;
QY 192 ELCAARVSEVLONRVRTEVGHVDFYASYYDLAAGVGLI 233
DB 5 EPCYAEVLVRVGRKLNQPEEVQGSFYASYYDRAVPTDMI 46
```

Search completed: June 8, 2004, 10:08:37
Job time : 60 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 10:10:12 ; Search time 48 Seconds
(without alignments)
1887.314 Million cell updates/sec

Title: US-09-905-589a-2_COPY_135_456
Perfect score: 1659
Sequence: 1 AKQIPDFMKATPLVXKAT.....ALGAIPIHYIDLNKQSPAS 322

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 830003

Minimum DB seq length: 0
Maximum DB seq length: 322

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTIS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	237	11.3	278	US-10-425-114-71791	Sequence 71791, A
2	184.5	11.1	233	US-10-425-114-39418	Sequence 39418, A
3	175.5	10.6	199	US-10-424-599-219735	Sequence 219735, A
4	168.5	10.2	106	US-10-424-599-20950	Sequence 20950, A
5	157.5	9.5	288	US-10-424-599-262339	Sequence 262339, A
6	151.5	9.1	206	US-09-833-245-870	Sequence 870, App
7	144.5	8.7	219	US-10-424-599-266621	Sequence 266621, A
8	126.5	7.5	119	US-10-424-599-160616	Sequence 160616, A
9	124.5	7.6	171	US-10-425-114-69638	Sequence 49638, A
10	120.5	7.3	135	US-10-425-114-52324	Sequence 52324, A
11	113	6.8	186	US-10-424-599-193675	Sequence 193675, A
12	110.5	6.7	141	US-10-425-114-43552	Sequence 43552, A
13	110.5	6.7	155	US-10-424-599-150130	Sequence 150130, A
14	109.5	6.6	220	US-10-002-631C-276	Sequence 276, App
15	104	6.3	46	US-09-864-761-34013	Sequence 34013, A

16	95.5	5.8	54	US-10-424-599-260917	Sequence 260917, A
17	93	5.6	157	US-10-424-599-143499	Sequence 143499, A
18	86	5.2	280	US-10-156-761-9488	Sequence 9488, App
19	85.5	5.2	236	US-10-424-599-227975	Sequence 227975, A
20	84	5.1	293	US-10-369-493-2970	Sequence 2970, App
21	83.5	5.0	224	US-10-156-761-12887	Sequence 12887, A
22	83	5.0	167	US-10-424-599-281704	Sequence 281704, A
23	83	5.0	290	US-10-287-226-280	Sequence 280, App
24	82.5	5.0	192	US-10-363-616-469	Sequence 469, App
25	82.5	5.0	138	US-10-264-049-2347	Sequence 2347, App
26	82.5	5.0	228	US-10-425-114-39004	Sequence 39004, A
27	82	4.9	158	US-10-374-780A-1409	Sequence 1409, App
28	81.5	4.9	254	US-10-403-105-8	Sequence 8, App11
29	81.5	4.9	299	US-10-104-047-2792	Sequence 158177, A
30	81	4.9	150	US-10-424-599-158177	Sequence 9481, App
31	80	4.8	280	US-10-156-761-9491	Sequence 5, App11
32	79.5	4.8	274	US-10-336-324-5	Sequence 36727, A
33	77	4.6	301	US-10-425-114-36727	Sequence 14007, A
34	76.5	4.6	209	US-10-369-493-14007	Sequence 146, App
35	76.5	4.6	252	US-09-984-130-146	Sequence 146, App
36	76.5	4.6	252	US-09-836-353A-146	Sequence 61268, A
37	76.5	4.6	287	US-10-282-122A-61268	Sequence 3, App11
38	76	4.6	210	US-10-288-986-3	Sequence 21969, App
39	76	4.6	200	US-10-428-599-211969	Sequence 477, App
40	76	4.6	273	US-10-389-566-477	Sequence 4735, App
41	75.5	4.6	215	US-10-369-493-4676	Sequence 287, App
42	75.5	4.6	215	US-10-369-493-7435	Sequence 287, App
43	75.5	4.6	286	US-09-731-872-287	Sequence 69397, A
44	75.5	4.6	286	US-09-876-997-287	
45	75	4.5	244	US-10-425-114-65997	

ALIGNMENTS

RESULT 1
US-10-425-114-71791
; Sequence 71791, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 71791
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Zea mays subsp. mexicana
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMROTEOSINTE009H08_FLI.pep
; US-10-425-114-71791
Query Match 14.3%; Score 237; DB 12; Length 278;
Best Local Similarity 25.8%; Pred. No. 2,5e-16;
Matches 73; Conservative 51; Mismatches 109; Indels 50; Gaps 11;
QY 68 ITTNFTGSLKTPGGSSVGMKLDGGSTGIATLPVREGTLQASPPG-----YLTALM 120
DB 1 VALNYYLIDKLGDPYSTAVVLDWGGSSVMAV--ANSANAASAPAPPHGEDPYVREYL 58
QY 121 FNTYLYGYSYSGLGIMAPLAIIGVGVGQAKDKEVSPCLSPSPFGMEHAEVTR 180
DB 59 KGDVNIYTHSYIHAFAPASRAEIL-----AKQGP--FSSCKLRGFGQTYNEQVD 110
QY 181 VSGQKAASLHEICAAVSEVLQNRVHREBKVD-----YAFSYVD 225

Db 111 ATAPEGA-VYKCREBEGAL--NLNAPQCMENCTFNWGGGAGQGGSIYVASSFYF 167
Qy 226 LAAGVGLIDAE-KGGSLVGDFFIAAKYVCRITLTOPO-----SSPFCMDLTY-V 274
Db 168 VASEVGVVDGNAPSGNTPGAFGSAEAKQMSVEEAKIEYPNVNDVDFYLCMDLAYQY 227
Qy 275 SLLOBEFGPRSKVLKTRKIDN---VETSWALGAIFHYIDS 313
Db 228 TLLVDGEGVETKEITVVDKVKGEYVEAANPLGSAIEAVSS 270

RESULT 2
US-10-425-114-39418
; Sequence 39418, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 39418
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700555327_FLI.pep
US-10-425-114-39418

Query Match 11.1%; Score 184.5; DB 12; Length 233;
Best Local Similarity 27.5%; Pred. No. 7.3e-11;
Matches 64; Conservative 33; Mismatches 91; Indels 45; Gaps 8;

Qy 114 YLTALRMFNRTYKLYSYLGLGMSARLAILGGVEQPAKDGKELVSPCLSPFSGEWE 173
Db 4 YIKTLVNGKTYDIYVSHLHFGKEASRAEML-KVTGDSA-----NPCILAGYNGTYT 55
Qy 174 HAEVTVYRSGQAAASLHELCAARVSEVLQNRVHRTVEEVGVD-----216
Db 56 YSGVKY-----KALASTSGSNFDCREVALKALKVNEPCPHONCTFGIWNNGGGSGQKV 110
Qy 217 FYAPSVYVYDLAAGVGLIDAEKGGSLV-GDFEIAAKYVCRITLTOPOSS-----PF 266
Db 111 LYVTSFYLVIOAGIADAKTSKYVPAEFKAAKACQVKEFDAQSTYPLMMDALPY 170
Qy 267 SCMDLTY-VSLLQEFQPRSKVLKTRKIDN---VETSWALGAIFHYIDS 314
Db 171 ICMDDITYQLTLVDGFLDPEWKEIIVANEIEYQCALVEGAWPLGSAIEAISL 223

RESULT 3
US-10-424-599-219735
; Sequence 219735, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 219735
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_40449C.1.pep
US-10-424-599-219735

Query Match 10.6%; Score 175.5; DB 12; Length 199;
Best Local Similarity 34.9%; Pred. No. 5.2e-10;
Matches 53; Conservative 21; Mismatches 59; Indels 19; Gaps 5;

Qy 1 AKQDIPDFWKATPLVLKATAGLRLPGEKAQKLQKVEFK-ASPFLVGDDCVSIMNG 59
Db 53 AQKQIPKTSKHTTSILFLYATAGVRRLLPDDDSKMLLDNAWSFLKSSSPFVCKRDVWKIISG 112
Qy 60 TDEGVSAWITINFLTGSIKT-PGGSSVGMGLDLGGSGTQIAFLPRVEGT-----LQ 108
Db 113 TEAYLGMIALNYDSGILGVKPKETTYGALDLGGSSLOVTF-----EGNRNKEQQNQQLLN 168
Qy 109 ASPFGYLTALRMFNRTYKLYSYLGLGMSA 140
Db 169 SETRLY---VRIGSMNHHLTGYSLAGYGLNEA 197

RESULT 4
US-10-424-599-209550
; Sequence 209550, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 209550
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_31250C.1.pep
US-10-424-599-209550

Query Match 10.2%; Score 168.5; DB 12; Length 106;
Best Local Similarity 38.5%; Pred. No. 1.1e-09;
Matches 35; Conservative 22; Mismatches 33; Indels 1; Gaps 1;

Qy 1 AKQDIPDFWKATPLVLKATAGLRLPGEKAQKLQKVEFKASPFILVG-DGVSIMNG 59
Db 16 AHSVVPREFRPKTPVRVGATAGLRALEGDASGRILQAVRDLLKORSTLKSEPDVAVLDG 75
Qy 60 TDEGVSAWITINFLTGSIKTPEGSSVGMGLDL 90
Db 76 TQEGAFQWVTNYLLGKLGKDFSETVGVDL 106

RESULT 5
US-10-424-599-262339
; Sequence 262339, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B

Page 3

Query Match	9.5%	Score 157.5	DB 12	Length 206
Best Local Similarity	42.3%	Pred. No. 4.5e-08		
Matches 33	Conservative 17	Mismatches 27	Indels 1	Gaps 1

```

60 TDEGVSAWITINFLTGSL 77
   ||| :||| :|
180 TQEGSYLWVTINYLLKXL 197

```

RE: PatentIn Ver. 2.1
NO 870

INFORMATION: Xaa equals any of the naturally occurring L-amino acids
245-870

```

QY      1 AKQDIPEDFKATPTPLVKATAQRLRLGSEKAQRL-GRKKEVFKASPFVLGGDDCVSIMG 59
      |  :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      155 AAAHVVKKKKETPLYLCTAGKRLLEPKQALAILADLVKDLPLEDFLISQSQAEVITG 211

```

```

60 TDEGVSAWITINFL-----TGSIXTPGSSVGMIDLGGSQIAF 99
      ||||| |||: ||||| |||: ||||| |||:
215 KQGVYAMIGINFLVLRFDHDESDAEATQELTAGRRRTVGILDMGAXXQIAY 26

```

Application US/10424595
Application No. US20040031072A1
INFORMATION.

INVENTOR: Cao Yongwei
 CO-INVENTOR: Soy Nucleic Acid Molecules and Other Molecules Associated With
 CO-INVENTOR: Plants and Uses Thereof for Plant Improvement
 EXPERIENCE: 38-21 (53223) B
 APPLICATION NUMBER: US/10/424, 599
 FILING DATE: 2003-04-28
 OF SEQ. ID NOS: 285684
 NO 286621

INFORMATION: Clone ID: PAT_MRT3847_82780C.1.paf
-599-266621

cal Similarity 25.9%; Pred. No. 1.2e-06;
57; Conservative 40; Mismatches 80; Indels 43; Gaps 10

```
1 MGLLRGWSVLLVGHGFVHGGLIAPAVLPISVAALSPRVLPPGAGSYNYGKSKFASS 60
```

```
61 : |||: | :||: : |||: : |||: : |||:  
GSSGASINECKSVLRALKVNESTCTHMKCTFGGIWNGGGDGQKTL--FVASFFFDRAA 118
```

```

119 EAGFADNLPVAIVRPADPEDAKQACQTKLENASTFPRVDEGNLPYLCMDLIYQYTL 178

```

```

179      : ||| : : : : : ||| : : |
      V D G F G I Y P M Q E I T L V K K Y D D A L V E A N P I G S A I E A V S S      218

```

ABSTRACT: Soy Nucleic Acid Molecules and Other Molecules Associated With
CDP INVENTION: Plants and Uses Thereof for Plant Improvement
REFERENCE: 38-21(53223)B
T APPLICATION NUMBER: US/10/424,599
T FILING DATE: 2003-04-28
OF SEQ ID NOS: 285684

```
; SEQ ID NO 160816
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_116234C.1.pep
US-10-424-599-160816

Query Match      7.6%; Score 126.5; DB 12; Length 119;
Best Local Similarity 27.8%; Pred. No. 0.00011;
Matches 27; Conservative 22; Mismatches 43; Indels 5; Gaps 1;

QY 40 EVKASPLVGDCCVIMWGTGSGVSAWITINFLTGLSKTPGSSVGMGLDILGGSGTQIAF 99
Db 1 QALRVSGFLFDMWARVISGEQISGWAVNALGNLGRPOETTGIVELGASLQV-- 58
QY 100 LPRVEGTQASPPGYLTALRMFNRTKLYSYSLGLG 136
Db 59 ---TSAKLNADIAQSLHTIRLSGWMYNLYTRSLPQLG 92

RESULT 9
US-10-425-114-49638
; Sequence 49638, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 49638
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700751504_FLI.pep
US-10-425-114-49638

Query Match      7.5%; Score 124.5; DB 12; Length 171;
Best Local Similarity 30.4%; Pred. No. 0.00011;
Matches 35; Conservative 21; Mismatches 42; Indels 17; Gaps 5;

QY 216 DFYAFSYYYDLAAGVGLIDAEKGGSLV-VGDFFEIAKYVCTLTETQPOSS----- 264
Db 48 NFFIASPFVEADBAEAGFVDNAPNAKVRPVDNFENAAKACNT-ELKDKLSIFPRVKGDV 106
QY 265 PFCMDLTY-VSLLLQERGFPRSKVLKLTTRKIDN---VETSWALGAIFHYDLSL 314
Db 107 PYICLDLVYETLLVDGFEIDPQOEITLVROVEYQDSLVAEAWPLGSAIEAISL 161

RESULT 10
US-10-425-114-52324
; Sequence 52324, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
```

```
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 52324
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700788139_FLI.pep
US-10-425-114-52324
```

```
Query Match      7.3%; Score 120.5; DB 12; Length 135;
Best Local Similarity 31.0%; Pred. No. 0.00021;
Matches 35; Conservative 21; Mismatches 42; Indels 15; Gaps 5;

QY 216 DFYAFSYYYDLAAGVGLIDAEKGGSLV-VGDFFEIAKYVCTLTETQPOSS-----P 265
Db 22 NLFVASPFEDRAABAGFADPNLPAIVIRPADFEAAKQACQTKLENAKSTFFRYDEGNLP 81
QY 266 FSCMDLTY-VSLLLQERGFPRSKVLKLTTRKIDN---VETSWALGAIFHYIDS 313
Db 82 YLCMDLIYQYTLVVDGFGIYFPWQETILVKKYYDDALVEAAWPLGSAIEAVSS 134
```

```
RESULT 11
US-10-424-599-195675
; Sequence 195675, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 195675
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(186)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_18720C.1.pep
US-10-424-599-195675
```

```
Query Match      6.8%; Score 113; DB 12; Length 186;
Best Local Similarity 25.7%; Pred. No. 0.0021;
Matches 49; Conservative 23; Mismatches 61; Indels 58; Gaps 10;

QY 148 VEGQPAKDGKEIV-----SPCLSPSPKGEWEAEVTVYRVSGQKAAASLHELCAARV 198
Db 20 ISGSSXDECREVVLOALKLINESCHONCTFGIWDGG----RGSGQK----- 62

QY 199 SEVLQNRVHRTVEEVKHVDYFAFSYYDIAAGVGLIDAEKGGSLV-VGDFFEIAKYVCT- 256
Db 63 -----ILFGTSSFFY-LPTBGTIGIIDLKPNKSIHPVDLBLEAKRACETK 105

QY 257 LE-----TOPOSS-----PFCMDLTY-VSLLLQERGFPRSKVLKLTTRKIDN---VETSWA 303
Db 106 LEDAKSTYPNTLAEDRLPVCLDIAIYQYALYTDGFLDPWQITVANELEYQDALVEAAWP 165

QY 304 LGAIFHYIDSL 314
Db 166 LGTAIEAISL 176
```

RESULT 12


```
; ORGANISM: Gallus gallus
US-09-240-639-16

Query Match      7.3%; Score 121; DB 4; Length 150;
Best Local Similarity 41.3%; Pred. No. 4.8e-05;
Matches 31; Conservative 10; Mismatches 30; Indels 4; Gaps 2;

QY 1 AKQDIPDFWKATPLVLKATAGLRLPCE---PGEKAQKLQKVEVFKASPFVLVGGDDCVSINMGTD 58
DB 66 AMEDVPKEKHADTPVYLGATAGMELLTIADPPSQTCLSAVMATLKSYPDFGG--AKILS 123

QY 59 GTDEGVSAWITINFL 73
DB 124 GEEGVFGWITANYL 138

RESULT 3
US-09-240-639-13
; Sequence 13, Application US/09240639
; Patent No. 6350447
; GENERAL INFORMATION:
; APPLICANT: Chadwick, Brian Paul
; APPLICANT: Frischauf, Anna-Maria
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO CD39-LIKE
; TITLE OF INVENTION: POLYPEPTIDES AND NUCLEIC ACIDS
; FILE REFERENCE: 9598-066
; CURRENT APPLICATION NUMBER: US/09/240,639
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-240-639-13

Query Match      6.7%; Score 111.5; DB 4; Length 153;
Best Local Similarity 40.0%; Pred. No. 0.00052;
Matches 34; Conservative 14; Mismatches 28; Indels 9; Gaps 4;

QY 1 AKQDIPDFWKATPLVLKATAGLRLPCEKAQKLQKVEVFKAS----PFLVGGDDCVSI 56
DB 65 AREVIPSQHQETPVYLGATAGMELLTME--SESLADRVLDVWERSLSNYFF--DFQGARI 121

QY 57 MNGTDEGVSAWITINFLTG--SLKT 79
DB 122 ITGQEGAYGWITINYLLGKFSQKT 146

RESULT 4
US-09-240-639-14
; Sequence 14, Application US/09240639
; Patent No. 6350447
; GENERAL INFORMATION:
; APPLICANT: Chadwick, Brian Paul
; APPLICANT: Frischauf, Anna-Maria
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO CD39-LIKE
; TITLE OF INVENTION: POLYPEPTIDES AND NUCLEIC ACIDS
; FILE REFERENCE: 9598-066
; CURRENT APPLICATION NUMBER: US/09/240,639
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-240-639-14

Query Match      6.7%; Score 110.5; DB 4; Length 154;
Best Local Similarity 40.5%; Pred. No. 0.00068;
Matches 30; Conservative 8; Mismatches 31; Indels 5; Gaps 2;
```

```
QY 5 IPDFWKATPLVLKATAGLRLPCE---KAQKLQKVEVFKASPFVLVGGDDCVSINMGTD 61
DB 70 IPASKQHQETPVYLGATAGMELLTME--PGEKAQKLQKVEVFKASPFVLVGGDDCVSINMGTD 127

QY 62 EGVSANWITINFLTG 75
DB 128 EGAYGWITINFLTG 141

RESULT 5
US-09-240-639-15
; Sequence 15, Application US/09240639
; Patent No. 6350447
; GENERAL INFORMATION:
; APPLICANT: Chadwick, Brian Paul
; APPLICANT: Frischauf, Anna-Maria
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO CD39-LIKE
; TITLE OF INVENTION: POLYPEPTIDES AND NUCLEIC ACIDS
; FILE REFERENCE: 9598-066
; CURRENT APPLICATION NUMBER: US/09/240,639
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-240-639-15

Query Match      6.6%; Score 109.5; DB 4; Length 153;
Best Local Similarity 38.5%; Pred. No. 0.00086;
Matches 30; Conservative 10; Mismatches 29; Indels 9; Gaps 3;

QY 1 AKQDIPDFWKATPLVLKATAGLRLPCE---PGEKAQKLQKVEVFKASPFVLVGGDDCVS 55
DB 65 ALQDVPKEXHAGTPLYLGATAGMELLTME--PGEKAQKLQKVEVFKASPFVLVGGDDCVS 120

QY 56 IMNGTDEGVSAWITINFL 73
DB 121 ILSQEGVFGWITANYL 138

RESULT 6
US-09-196-281-8
; Sequence 8, Application US/09196281A
; Patent No. 6605458
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter K.
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; TITLE OF INVENTION: Protease Variants And Compositions
; FILE REFERENCE: 5435.200-US
; CURRENT APPLICATION NUMBER: US/09/196,281A
; CURRENT FILING DATE: 1998-11-19
; EARLIER APPLICATION NUMBER: 1332/97
; EARLIER FILING DATE: 1997-11-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Bacillus
US-09-196-281-8

Query Match      4.9%; Score 81.5; DB 4; Length 274;
Best Local Similarity 22.0%; Pred. No. 2.3;
Matches 69; Conservative 50; Mismatches 118; Indels 77; Gaps 17;

QY 22 GLRLPGEKAQKLQKVEVFKASPFVLVGGDDCVSINMGTDGVSAWITINFLTGLSKTPG 81
DB 7 GIPLIKADKVG-----AQGYKANVKGI-----IDTGIAASVEAAAQHTDLKVG 52

QY 82 GSSVGM-----LDLGGGSGTQ-----IAFLPRVEGTQLQSPFGLTALRMFNRTYKLSYSY 132
```

Db 53 GASFSVGSNTYDNGHGHVAGTVAALDNTTGVGVAPNLSVAIKLNS----- 104
QY 133 LGLGMSARLAILGVGEPAPKDKELVSPCL-SPSPFGMEHAEVTVRVSQKAAALH 191
Db 105 -GSGTYS---ALVSIIE-WATONGLDVIMSLGSPS-----GSTALKQAVDKAYASGI 152
QY 192 ELCAA--RVSEVLQNRVHRTVEVHVDPYAFSYTYDLAAGVGLDAEKG--SLVVG 245
Db 153 VVVAAGNSGSSGQNTI-----GYPKYDSVIAGVADSNKKNASFSVGA 200
QY 246 FEIAKYCTRLTEPOSPSPSCMDLTYY-----SLLQFEPKRSKVLTLTKID 296
Db 201 LEVAPGV-SVYSTYP-SNTYTSLNGTSMASPHVAGAAALITSK--YPTLSASQVNRNLS 256
QY 297 NVETSMALGALFHY 310
Db 257 STATN-IGDSFY 268

RESULT 7
US-08-264-534-4
Sequence 4, Application US/08264534
Patent No. 5648464
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon et al.
TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains
TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/264,534
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/695,189
FILING DATE: 03-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-004
TELEPHONE: 212 790-9090
TELEFAX: 212 869864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 199 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-264-534-4

Query Match 4.9%; Score 81; DB 1; Length 199;
Best Local Similarity 21.6%; Pred. No. 1.6;
Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;
QY 14 PLYKATAGRLIPGKAKQLQKVEVKASPLVGVDDCVSINAGTDEGVSAWTTINFL 73
Db 26 PAELRATKTIKGSPPCTTAFRL-----CLKEVOTTEGASISTGCSFG 67

QY 74 TGSILTPGSSVGMULDGGSTQIAFLPVEGTLOASPPGYLTALRMNRTKLYSYSL 133
Db 68 NATTKILGSSSFLSPGVGATVLPFTFWTYSFTL-----ILQALDMNTSY----- 115
QY 134 GLGMSARLAILGVGEPAPKDKELVSPCLSPSPFG-----EME-----HAETVR 180
Db 116 -----PDAERLIE---ETSISGVILPSPFKTLDHTRNRNRIYR 152
QY 181 VSGQKAAASLHEIC 194
Db 153 VRVQCAVITYNTTC 166

RESULT 8
US-08-465-500-4
Sequence 4, Application US/08465500
Patent No. 5789195
GENERAL INFORMATION:
APPLICANT: Muskavitch, Marc A.T.
APPLICANT: Fenon, Richard G.
APPLICANT: Reday, Iarla
APPLICANT: Blumweller, Christine M.
APPLICANT: Shepard, Scott B.
TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS
TITLE OF INVENTION: IN TOPORYTHMIC PROTEINS, AND METHODS BASED THEREON
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,500
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-034
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8664/9741
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 199 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-465-500-4

Query Match 4.9%; Score 81; DB 1; Length 199;
Best Local Similarity 21.6%; Pred. No. 1.6;
Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;
QY 14 PLYKATAGRLIPGKAKQLQKVEVKASPLVGVDDCVSINAGTDEGVSAWTTINFL 73
Db 26 PAELRATKTIKGSPPCTTAFRL-----CLKEVOTTEGASISTGCSFG 67
QY 74 TGSILTPGSSVGMULDGGSTQIAFLPVEGTLOASPPGYLTALRMNRTKLYSYSL 133
Db 68 NATTKILGSSSFLSPGVGATVLPFTFWTYSFTL-----ILQALDMNTSY----- 115
QY 134 GLGMSARLAILGVGEPAPKDKELVSPCLSPSPFG-----EME-----HAETVR 180

Db 116 -----PDAERLIE---ETSYSVILPSPWKTLDHIGNARITYR 152

QY 181 VSGKAAASLHELIC 194

Db 153 VRVQCAVYYNTTC 166

RESULT 9

US-08-346-126-4

; Sequence 4, Application US/08346126

; Patent No. 5849869

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon et al.

; TITLE OF INVENTION: Human No. 5849869ch And Delta, Binding Domains

; TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/346,126

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/791,923

; FILING DATE: 14-NOV-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Mirock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-007

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 790-9090

; TELEFAX: 212 8698864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 199 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: unknown

; MOLECULE TYPE: peptide

US-08-346-126-4

Query Match 4.9%; Score 81; DB 2; Length 199;

Best Local Similarity 21.6%; Pred. No. 1.6;

Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;

QY 14 PLVLKATAGLRLPGEKAQKLLQKVEFKASPLVGGDDCVSIMNGTDEGVSAWITINFL 73

Db 26 PAELRATKTCSPCTTAFL-----CLKEYQTTEQGASISTGCSFG 67

QY 74 TGSLLKTPGSSVGMLDLGGSTQIAFLPRVEGTLOASPPGYLTALRMFNRTYKLYSYSL 133

Db 68 NATTKILGSSFVLSDPGVGAIVLPFTFRWTKSFTL-----ILQALDMYNTSY----- 115

QY 134 GLGLMSARLAILGGVEGQPAKGKELVSPCLSPSPFKG-----EWE-----HAEVYR 180

Db 116 -----PDAERLIE---ETSYSVILPSPWKTLDHIGNARITYR 152

QY 181 VSGKAAASLHELIC 194

Db 153 VRVQCAVYYNTTC 166

RESULT 11

US-08-893-828-4

; Sequence 4, Application US/08893828

; Patent No. 6050922

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Muskavitch, Marc A.T.

RESULT 10

US-08-346-128-4

; Sequence 4, Application US/08346128

; Patent No. 5856441

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon et al.

; TITLE OF INVENTION: Human No. 5856441ch And Delta, Binding Domains

; TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon

; NUMBER OF SEQUENCES: 37

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/346,128

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/879,038

; FILING DATE: 30-APR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Mirock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-009

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 790-9090

; TELEFAX: 212 8698864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 199 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: unknown

; MOLECULE TYPE: peptide

US-08-346-128-4

Query Match 4.9%; Score 81; DB 2; Length 199;

Best Local Similarity 21.6%; Pred. No. 1.6;

Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;

QY 14 PLVLKATAGLRLPGEKAQKLLQKVEFKASPLVGGDDCVSIMNGTDEGVSAWITINFL 73

Db 26 PAELRATKTCSPCTTAFL-----CLKEYQTTEQGASISTGCSFG 67

QY 74 TGSLLKTPGSSVGMLDLGGSTQIAFLPRVEGTLOASPPGYLTALRMFNRTYKLYSYSL 133

Db 68 NATTKILGSSFVLSDPGVGAIVLPFTFRWTKSFTL-----ILQALDMYNTSY----- 115

QY 134 GLGLMSARLAILGGVEGQPAKGKELVSPCLSPSPFKG-----EWE-----HAEVYR 180

Db 116 -----PDAERLIE---ETSYSVILPSPWKTLDHIGNARITYR 152

QY 181 VSGKAAASLHELIC 194

Db 153 VRVQCAVYYNTTC 166

RESULT 11

US-08-893-828-4

; Sequence 4, Application US/08893828

; Patent No. 6050922

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Muskavitch, Marc A.T.

APPLICANT: Rehon, Richard G.
APPLICANT: Blumauer, Christine M.
APPLICANT: Shepard, Scott B.
TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS
TITLE OF INVENTION: IN TOPOGRAPHIC PROTEINS, AND METHODS BASED THEREON
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSER: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/893,828
FILING DATE: 11-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-050
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 199 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-893-828-4

Query Match 4.9%; Score 81; DB 3; Length 199;
Best Local Similarity 21.6%; Pred. No. 1.6;
Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;

QY 14 PLVLTATGRLPLPEKAKLQKVEYKASPLVGDCCVIMNGDEGVSAWITINFL 73
DB 26 PAELRATITIGSPCTTAFRL-----CLKEYQTTEGASISTGCSFG 67
QY 74 TGSLLTPGSSSVGMIDLGGSTQIAFLPRVEGTLOASPPGYLTALMFMNRYKLYSYSL 133
DB 68 NATIKLIGSSSVLSDPGVAIVLPFTFRWTKSFTL-----ILQALDMVNTSY----- 115
QY 134 GLGMSARLAILGVGEGPAKDGKELVSPCLSPFGK-----EWE-----HAEVYR 180
DB 116 -----PDAERLIE-----ETSYSGVILPSPWKTLDHGRNARITR 152
QY 181 VSGOKAASLHSLC 194
DB 153 VRVQCAVITYNTTC 166

RESULT 12
US-08-264-534-9
Sequence 9, Application US/08264534
Patent No. 5648464
GENERAL INFORMATION:
APPLICANT: Aravanis-Tsakonas, Spyridon et al.
TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains
TITLE OF INVENTION: In Topographic Proteins, And Methods Based Thereon
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSES: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York

STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/264,534
FILING DATE: 03-MAY-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/695,189
FILING DATE: 03-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 293 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-264-534-9

Query Match 4.9%; Score 81; DB 1; Length 293;
Best Local Similarity 21.6%; Pred. No. 2.9;
Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;

QY 14 PLVLTATGRLPLPEKAKLQKVEYKASPLVGDCCVIMNGDEGVSAWITINFL 73
DB 106 PAELRATITIGSPCTTAFRL-----CLKEYQTTEGASISTGCSFG 147
QY 74 TGSLLTPGSSSVGMIDLGGSTQIAFLPRVEGTLOASPPGYLTALMFMNRYKLYSYSL 133
DB 148 NATIKLIGSSSVLSDPGVAIVLPFTFRWTKSFTL-----ILQALDMVNTSY----- 195
QY 134 GLGMSARLAILGVGEGPAKDGKELVSPCLSPFGK-----EWE-----HAEVYR 180
DB 196 -----PDAERLIE-----ETSYSGVILPSPWKTLDHGRNARITR 232
QY 181 VSGOKAASLHSLC 194
DB 233 VRVQCAVITYNTTC 246

RESULT 13
US-08-083-590A-4
Sequence 4, Application US/08083590A
Patent No. 5786158
GENERAL INFORMATION:
APPLICANT: Aravanis-Tsakonas, S. et al.
TITLE OF INVENTION: Therapeutic And Diagnostic Methods
TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSES: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS

```
/
/
/   SOFTWARE: PatentIn Release #1.0, Version #1.25
/   CURRENT APPLICATION DATA:
/   APPLICATION NUMBER: US/08/083,590A
/   FILING DATE: 25-JUN-1993
/   CLASSIFICATION: 435
/   ATTORNEY/AGENT INFORMATION:
/   NAME: Mierock, S. Leslie
/   REGISTRATION NUMBER: 18,872
/   REFERENCE/DOCKET NUMBER: 7326-015
/   TELECOMMUNICATION INFORMATION:
/   TELEPHONE: 212 790-9090
/   TELEFAX: 212 8698864/9741
/   TELEX: 66141 PENNIE
/   INFORMATION FOR SEQ ID NO: 4:
/   SEQUENCE CHARACTERISTICS:
/   LENGTH: 293 amino acids
/   TYPE: amino acid
/   TOPOLOGY: unknown
/   MOLECULE TYPE: protein
/   US-08-083-590A-4
/
/   Query Match          4.9%; Score 81; DB 1; Length 293;
/   Best Local Similarity 21.6%; Pred. No. 2.9;
/   Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;
/
/   QY 14 PLVLKATAGRLRLPGKAKLQKVEFKASFFLVGGDDCVSIMNGTDEGVSAMITINFL 73
/   Db 106 PAELRATKTIKGCSPCTTAFRL-----CLKEYQTTEQGASISTGCSFG 147
/
/   QY 74 TGS�KTPGSSVGMGLDGGSTQIAFLPRVEGTLOASPPGYLTALRMFNRTYKLYSYL 133
/   Db 148 NATTKILGGSSFFVLSDPGVAIVLPFTFRWTKSFLL-----ILOALDMNTSY----- 195
/
/   QY 134 GLGLMSARLAILGGVEGQPAKGKELVSPCLSPSPFKG-----EWE-----HAEVTVYR 180
/   Db 196 -----PDAERLIE-----ETSYSGVILPSPENKTLDHIGNRARIYR 232
/
/   QY 181 VSCQKAAASLHELIC 194
/   Db 233 VRVQCAVYYNTTC 246
/
/   RESULT 14
/   US-08-465-500-9
/   ; Sequence 9, Application US/08465500
/   ; Patent No. 5789195
/   ; GENERAL INFORMATION:
/   ; APPLICANT: Artavanis-Tsakonas, Spyridon
/   ; APPLICANT: Muskavitch, Marc A.T.
/   ; APPLICANT: Feron, Richard G.
/   ; APPLICANT: Rebay, Ilaria
/   ; APPLICANT: Blummueller, Cristine M.
/   ; APPLICANT: Shepard, Scott B.
/   ; TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS
/   ; TITLE OF INVENTION: IN TOPORYTHMIC PROTEINS, AND METHODS BASED THEREON
/   ; NUMBER OF SEQUENCES: 34
/   ; CORRESPONDENCE ADDRESS:
/   ; ADDRESSEE: PENNIE & EDMONDS
/   ; STREET: 1155 Avenue of the Americas
/   ; CITY: New York
/   ; STATE: NY
/   ; COUNTRY: USA
/   ; ZIP: 10036-2711
/   ; COMPUTER READABLE FORM:
/   ; MEDIUM TYPE: Floppy disk
/   ; COMPUTER: IBM PC compatible
/   ; OPERATING SYSTEM: PC-DOS/MS-DOS
/   ; SOFTWARE: PatentIn Release #1.0, Version #1.30
/   ; CURRENT APPLICATION DATA:
/   ; APPLICATION NUMBER: US/08/465,500
/   ; FILING DATE: 05-JUN-1995
/   ; CLASSIFICATION: 435
/   ; ATTORNEY/AGENT INFORMATION:
```

```
/
/
/   NAME: Mierock, S. Leslie
/   REGISTRATION NUMBER: 18,872
/   REFERENCE/DOCKET NUMBER: 7326-034
/   TELECOMMUNICATION INFORMATION:
/   TELEPHONE: (212) 790-9090
/   TELEFAX: (212) 869-8864/9741
/   INFORMATION FOR SEQ ID NO: 9:
/   SEQUENCE CHARACTERISTICS:
/   LENGTH: 293 amino acids
/   TYPE: amino acid
/   TOPOLOGY: unknown
/   MOLECULE TYPE: protein
/   US-08-465-500-9
/
/   Query Match          4.9%; Score 81; DB 1; Length 293;
/   Best Local Similarity 21.6%; Pred. No. 2.9;
/   Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;
/
/   QY 14 PLVLKATAGRLRLPGKAKLQKVEFKASFFLVGGDDCVSIMNGTDEGVSAMITINFL 73
/   Db 106 PAELRATKTIKGCSPCTTAFRL-----CLKEYQTTEQGASISTGCSFG 147
/
/   QY 74 TGS�KTPGSSVGMGLDGGSTQIAFLPRVEGTLOASPPGYLTALRMFNRTYKLYSYL 133
/   Db 148 NATTKILGGSSFFVLSDPGVAIVLPFTFRWTKSFLL-----ILOALDMNTSY----- 195
/
/   QY 134 GLGLMSARLAILGGVEGQPAKGKELVSPCLSPSPFKG-----EWE-----HAEVTVYR 180
/   Db 196 -----PDAERLIE-----ETSYSGVILPSPENKTLDHIGNRARIYR 232
/
/   QY 181 VSCQKAAASLHELIC 194
/   Db 233 VRVQCAVYYNTTC 246
/
/   RESULT 15
/   US-08-346-126-9
/   ; Sequence 9, Application US/08346126
/   ; Patent No. 5849869
/   ; GENERAL INFORMATION:
/   ; APPLICANT: Artavanis-Tsakonas, Spyridon et al.
/   ; TITLE OF INVENTION: Human No. 5849869ch And Delta, Binding Domains
/   ; TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon
/   ; NUMBER OF SEQUENCES: 30
/   ; CORRESPONDENCE ADDRESS:
/   ; ADDRESSEE: Pennie & Edmonds
/   ; STREET: 1155 Avenue of the Americas
/   ; CITY: New York
/   ; STATE: New York
/   ; COUNTRY: U.S.A.
/   ; ZIP: 10036
/   ; COMPUTER READABLE FORM:
/   ; MEDIUM TYPE: Floppy disk
/   ; COMPUTER: IBM PC compatible
/   ; OPERATING SYSTEM: PC-DOS/MS-DOS
/   ; SOFTWARE: PatentIn Release #1.0, Version #1.25
/   ; CURRENT APPLICATION DATA:
/   ; APPLICATION NUMBER: US/08/346,126
/   ; FILING DATE:
/   ; CLASSIFICATION: 530
/   ; PRIOR APPLICATION DATA:
/   ; APPLICATION NUMBER: 07/791,923
/   ; FILING DATE: 14-NOV-1991
/   ; ATTORNEY/AGENT INFORMATION:
/   ; NAME: Mierock, S. Leslie
/   ; REGISTRATION NUMBER: 18,872
/   ; REFERENCE/DOCKET NUMBER: 7326-007
/   ; TELECOMMUNICATION INFORMATION:
/   ; TELEPHONE: 212 790-9090
/   ; TELEFAX: 212 8698864/9741
/   ; TELEX: 66141 PENNIE
/   ; INFORMATION FOR SEQ ID NO: 9:
/   ; SEQUENCE CHARACTERISTICS:
```

LENGTH: 293 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-346-126-9

Query Match 4.9%; Score 81; DB 2; Length 293;
Best Local Similarity 21.6%; Pred. No. 2.9;
Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;

QY	14	PLVKAATAGADLLPGEKAKQKLLQKVEVFKASPFVGGDDCVSNMGTDGVSAMITINFL	73
DB	106	PAELPATKTIICSPCTIHFRL-----CLKEYQTTEQASISITGCSFG	147
QY	74	TGSLKTPGGSSVGMLDLGGGSTQIAFLPRVEGTLQASPPGYLTALRMFNRTYKLYSYSYL	133
DB	148	NATTKILGSSSFVLDPGVGAIVLPFTFRWTKSFLL--ILQALDMYNTSY-----	195
QY	134	GLGMSARLAILGVEGQPAKDGKELVSPCLSPSPFKG-----EWE-----HAEVTVR	180
DB	196	-----PDAERLIE--ETISYGVILLPSPEWKTLDHIGRWARITTR	232
QY	181	VSGORAAASLHELIC	194
DB	233	VRVQCAVTYYNTTC	246

Search completed: June 8, 2004, 10:11:13
Job time : 24 secs